

Result No.	Score	Query Match	Length	DB	ID	Description
1	6042	76.5	1481	2	US-09-251-645-14	Sequence 14, Appl
2	6031.5	76.3	1476	2	US-09-817-514A-4	Sequence 4, Appl
3	4581.5	58.0	1485	2	US-08-851-567B-32	Sequence 32, Appl
4	1601.5	2.0	1426	2	US-09-492-709A-340	Sequence 340, Appl
5	155.5	2.0	658	2	US-09-252-991A-24910	Sequence 24910, A
6	155.5	2.0	2315	2	US-09-543-681A-5434	Sequence 5434, Ap
7	154	1.9	4630	2	US-09-091-609-2	Sequence 2, Appl
8	154	1.9	5215	2	US-09-105-537-2	Sequence 2, Appl
9	153.5	1.9	1028	2	US-09-543-681A-7181	Sequence 7181, Ap
10	144	1.8	1377	2	US-09-711-164-467	Sequence 467, App
11	143	1.8	798	2	US-09-489-039A-10045	Sequence 10045, A
12	142.5	1.8	1183	1	US-08-447-031A-2	Sequence 2, Appl
13	140	1.8	979	1	US-08-346-455B-38	Sequence 38, Appl
14	140	1.8	979	2	US-08-977-221-38	Sequence 38, Appl
15	140	1.8	979	2	US-09-483-831B-70	Sequence 70, Appl
16	140	1.8	979	4	PCF-US95-06613-38	Sequence 38, Appl
17	139.5	1.8	646	2	US-09-902-540-10353	Sequence 10353, A
18	139.5	1.8	1577	1	US-08-793-824-2	Sequence 2, Appl
19	139	1.8	2200	2	US-09-796-575-2	Sequence 2, Appl
20	138	1.7	1529	2	US-09-215-694-1	Sequence 1, Appl
21	138	1.7	1529	2	US-10-109-310-1	Sequence 1, Appl
22	137	1.7	2628	1	US-08-570-311-14	Sequence 14, Appl
23	136.5	1.7	990	1	US-08-232-540-2	Sequence 2, Appl
24	136.5	1.7	990	1	US-08-428-949A-2	Sequence 2, Appl
25	136.5	1.7	990	1	US-08-428-948A-2	Sequence 2, Appl
26	136.5	1.7	990	1	US-08-428-946-2	Sequence 2, Appl
27	136.5	1.7	990	4	PCF-US95-04656-2	Sequence 2, Appl

```
Db 301 YGFALTRRLCRQLMFHRLQTLGSGQAKGDBEPALVSRLLDYDENAVVSTLVSVERVGH 360
Qy 355 EQDG-NVVTLPPLLELAYQDPSPRHHAHQMDVLANFNAIQRWQLVDLKGEGPLGLLYOD 413
Db 361 EQDGTAVALPPLLELAYQDPEPEQKALRPMMDVLANFNITQRMQLDLQEGVGILYQD 420
Qy 414 KGAWVRSRQRLGEIGSDAVTWKMOPLSVIPSLQSNASLVINGDQOLDWVITGPGLAG 473
Db 421 KGAWVRSRQRLGEIGSDAVTWKMOPLSVIPSLQSNASLVINGDQOLDWVITGPGLAG 480
Qy 474 YHSQRPDGSGWTRFTPLNALPVEYTHPRAQLADLMGAGLSDLVLIGPKSVRLYANTRDGF 533
Db 481 YHSQHPDGSGWTRFTPLHALPIEYTHPRAQLADLMGAGLSDLVLIGPKSVRLYANNRDGT 540
Qy 534 KGADVQSGDITLVPFGADPRKLVAFSDVLGSGQAHLEVSATKVTCPWNLGRGRFGQPI 593
Db 541 EGRDVVQSGDITLVPFGADPRKLVAFSDVLGSGQAHLEVSATKVTCPWNLGRGRFGQPI 600
Qy 594 TLPFGSQPATEFNPAPVYLADLDGSGPTDLYVHTNRLDI FLKNGSGPAEPVTLRFPFG 653
Db 601 TLPFGSQSANFPDRVHLADLDGSGPADLIYVHADHLDFSNESNGPAQPTFLRFPDG 660
Qy 654 LRFDPHTCOLQADVQGLGVVSLILSVPHMSPHRCDLTNMKPMLNEMNNNGVHHTLR 713
Db 661 LRFDDTCQLQADVQGLGVVSLILSVPHMAPHRCDLTNAKPMLSEMNNNGAHTTLH 720
Qy 714 YRSSQFWDKKAALTTGQTPVCYLPFPPIHTLWQTEDEISGNKLVTLRLYARGAWDG 773
Db 721 YRSSQFWDKKAALATGQTPCYLPFPVHTLWQTEDEISGNKLVTLRLYARGAWDG 780
Qy 774 RERFEGFVGEQDTHQALQAGNAPERTPALTKNMYATGLPIDNALSYEWR-DDQAF 832
Db 781 RERFEGFVGEQDTHQALQAGNAPERTPALTKNMYATGLPIDNALSYEWR-DDQAF 840
Qy 833 AGFSRFTTWQDNKDVPLTPEDDNRSYWFNRALKGQLLSELVGLDSDSTNKHYRFTVTEF 892
Db 841 TGTPTPHTLWKEGKDVPLTPEDDNLNLYNLNALKGQPLRSELVGLDGSAAQKIPYVTES 900
Qy 893 RSQVRLQHTDSRPVLMSSVRSNHYHRIASDFCQSNITLSDRFGQPLKQLSVQY 952
Db 901 RPQVRQLQDNTLSPVLWASVRSYHYRIISDPQCNDITLSSDLFGQPLKQLSVQY 960
Qy 953 PRQCPALNLYPDLTKLLANSYDQOROLRTYQSSWHHLTNNTVRLGLPDPSTRSD 1012
Db 961 PRNKPTNYPDLTKLTFASSYDQQLRLTYQSSWHHLIANELRVLGLPDPGTRSD 1020
Qy 1013 IFTYGAENVPAAGLNLLELLSDKNSLIADDPREYLGQOKTAY----TDGQNTTFLQTPTR 1068
Db 1021 AFTYDAKHVPVQGLNLEALCAENSLIADDPREYLNQRTFYTDGKTDGKNPTPLKTPTR 1080
Qy 1069 QALIAETETTVFNQSTLSAFNGSI PSDKLSSTLEQAGYQNTNLYLPRTGBDKVWAHGY 1128
Db 1081 QALIAETETAVLTESLISAFDGGITPDELFGLLTQAGYQOEYLPFLSGENQVWARKGY 1140
Qy 1129 TDYGTAAQFWRPKQKSNLTQKITLIWDANYCVVQTRDAAGLTTSKDYDWEFLTPVOL 1188
Db 1141 TDYGTVEQFWRPVAQRNTQTKTLLKMDTHYCVITQTQDAAGLTVSANTDWRFLTPMQL 1200
Qy 1189 TDINDNQHLLTDLALGRPITLRFWGTENGQHTGYSSPEKASFPSPDVNAAILKXPLPV 1248
Db 1201 TDINDNVHIIITDLALGRPVTPQFWGIENGAVATGYSSEPAKFPPTPPVDVNAIALTGPLPV 1260
Qy 1249 AQCVVAPESWMPVLSQTNRLABQDWQKLKNARIITEDGRICTLAYRWVWSQKAIQ 1308
Db 1261 AQCLVAPDSWMPVLSQTEFTNTLTQBEQKTLRLRIITEDWRICALARRWLOSQAGTP 1320
Qy 1309 LISLLNNGPPLPHSLTLTDDRVDHDPQEQIROQVVFSDGFGRLQAAARHEAGMARQ 1368
Db 1321 LVKLLTNSIGLPPHNLMLATDRDRSEEQIROQVAFSDGFGRLQAAARHEAGMARQ 1380
Qy 1369 EDGSLINVOHTENRNVATGRTEYDNKGQPIRTYQPYFLNDWRVYVNSDARSQEKAYADT 1428
```

```
Db 1381 QDGLSVTKMEDTKTRWAITGRTEYDNKGQAI RTYQPYFLNDWRVYVNSDARS--KEAYADT 1438
Qy 1429 HYVDPTGREIKVITAKGWFRTLTFTPMFTVNEDENDTAAE 1468
Db 1439 HYIDPIGREIQVITAKGMLRQNRQYFPWFVTSEENDLSAD 1478
RESULT 2
US-09-817-514A-4
; Sequence 4, Application US/09817514A
; Patent No. 6639129
; GENERAL INFORMATION:
; APPLICANT: ffrench-Constant, Richard
; APPLICANT: Bowen, David
; APPLICANT: Rocheleau, Thomas
; APPLICANT: Waterfield, Nicholas
; TITLE OF INVENTION: DNA SEQUENCES FROM PHOTORHABDUS LUMINESCENS
; FILE REFERENCE: 61645
; CURRENT APPLICATION NUMBER: US/09/817,514A
; PRIOR FILING DATE: 2000-03-26
; PRIOR APPLICATION NUMBER: US 60/191806
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 1476
; TYPE: PRT
; ORGANISM: Photorhabdus luminescens
; US-09-817-514A-4
Query Match 76.3%; Score 6031.5; DB 2; Length 1476;
Best Local Similarity 75.9%; Pred. No. 0;
Matches 1120; Conservative 122; Mismatches 224; Indels 9; Gaps 3;
Qy 1 MQNSQDPSITELSLPKGGGAIITGMGEALPTPTGPDGMAALSLPLPISAGRGYAPAFITLNN 60
Db 1 MQNSQTFSTELSLPKGGGAIITGMGEALTPAGPDGMAALSLPLPISAGRGYAPSLTINN 60
Qy 61 SGAGNSPGLGWDGCVNMTIRRETHFGVPHYDSTDTPLGPEGEVLVVA-----DQPRDES 114
Db 61 SGTGNSPGLGWDGCVNMTIRRETHFGVPHYDSTDTPLGPEGEVLVVAALNAGQADIRSES 120
Qy 115 TLQGINLGAFTFTVYRSRLESFHSFSLYQWPKTKTDFWLIYSPDGQVHLLGKSPQAR 174
Db 121 SLQGINLGAFTFTVYRSRLESFHSFSLYQWPKTKTDFWLIYSPDGQVHLLGKSPQAR 180
Qy 175 ISNPSQTTTAQWLLEASVSSRGEIYYQYRAEDDPTGCEADEITHTLQATAQRYLHYVY 234
Db 181 ISNPLNVNQTALWLLLEASISHSSEQIYYQYRAEDEAGCETDELAHPSATVQRYLQTVHY 240
Qy 235 GNRTASETLPGLDGSGAPSOADMFLYLVFDYDGRSNNLKTTPAPSTTGSMLCRDRESRYE 294
Db 241 GNLTASDVFPPTLNGDDPLKSGWMFCLVFDYDGRSNNLKTTPAPSTTGSMLCRDRESRYE 300
Qy 295 YGFEIRTRLCROVLMYHHLQALDSKITENHNPPTLVSRILANYDESAIASTLVFVRVGH 354
Db 301 YGFELTRLCRQILMFHRLQTLGSGQAKGDBEPALVSRLLDYDENAVVSTLVSVERVGH 360
Qy 355 EODGNVVTLPPLLELAYQDPSPRHHAHQMDVLANFNAIQRWQLVDLKGEGPLGLLYQDK 414
Db 361 EDNNTVTALPPLLELAYQDPEPEQKALRPMMDVLANFNITQRMQLDLKGEVGPGLYQDR 420
Qy 415 GAWWVRSRQRLGEIGSDAVTWKMOPLSVIPSLQSNASLVINGDQOLDWVITGPGLAGY 474
Db 421 NGWVYSRQRLGEIGSDAVTWKMOPLSVIPSLQSNASLVINGDQOLDWVITGPGLAGY 480
Qy 475 HSQRPDGSGWTRFTPLNALPVEYTHPRAQLADLMGAGLSDLVLIGPKSVRLYANTRDGF 534
Db 481 HSQHPDGSGWTRFTPLHALPIEYTHPRAQLADLMGAGLSDLVLIGPKSVRLYANNRDGT 540
Qy 535 GKDVVQSGDITLVPFGADPRKLVAFSDVLGSGQAHLEVSATKVTCPWNLGRGRFGQPI 594
Db 541 GRDVVQSGDITLVPFGADPRKLVAFSDVLGSGQAHLEVSATKVTCPWNLGRGRFGQPI 600
```

```
Qy 595 L P G S Q P A T E N P A Q V Y L A D L G S G P T D L I Y V H T N R L D I F L N K S G N G P A E P V T L R F P P G L 654
Db 601 L P G S Q S A S F N P R V H L A D L G S G P A D L I Y V H A D R L D I F S N E S G N G P A K P F T L S F P D G L 660
Qy 655 R F D H T C Q L Q M A D V Q L G V A S I L S V P H M S P H H W R C D L T N M K P W L L N E M N N M G V H H T L A R Y 714
Db 661 R F D T C Q L Q M A D V Q L G V A S I L S V P H M S P H H W R C D L T N A K P W L L S E T N N M G N H T L H Y 720
Qy 715 R S S Q F W L D E K A A A L T T G Q T P V C Y L P P I P H T L M O T E T E D E I S G N K L V T L R Y A R G A M D G R 774
Db 721 R S S V Q F W L D E K A A A L T G Q T P V C Y L P P P V H T L M O T E T E D E I S G N K L V T L R A Y A H G A M D G R 780
Qy 775 E R E R G F G Y V E Q T S H Q L A Q N A P E R T P P A L T K W Y A T G L P I D N A L S T E W R - D D Q A F A 833
Db 781 E R E R G F G Y V E Q T S H Q L A Q N A P E R T P P A L T K W Y A T G L P A V D N A L S A G Y W R G D K Q A F A 840
Qy 834 G F S P R F T T W O N K D V P L T P E D N S R Y F N R A L K G O L L R S E L Y G L D D S T N K H V P V T T E F R 893
Db 841 G F T P R F T L W K G K D V P L T P E D D H N L Y W L N R A L K G O P L R S E L Y G L D S A Q Q O I P T V T T E S R 900
Qy 894 S Q V R L Q H T D S R Y P V L M S V V E S R N Y H Y E R I A S D P Q C S Q N I T L S S D R F G Q P L K Q L S V Q Y P 953
Db 901 P O V R Q L O D G A T V S P V L M A S V V E S R Y H Y E R I I S D P Q C N Q D I T L S S D L F G Q P L K Q V S V Q Y P 960
Qy 954 R R O Q P A I N L Y P D T L P D K L L A N S Y D D Q Q R L R T Y Q Q S S W H L T N N T V R V L G L P D S T R S D I 1013
Db 961 R R N E P T T P Y P D T L P D T L F A S S Y D D Q Q L L R L T C R Q S S W H L I G N E L R V L G L P D G T R S D A 1020
Qy 1014 F T Y G A E N P A G L N L E L L S D K N S I A D D K P R E Y L G Q O K T A Y T D G O N T P L O T P T R Q A L I A 1073
Db 1021 F T Y D A K V P D G L M L E T I C A E N S I A D D K P R E Y L N Q O R F T Y T D G N Q O T P L K T P T R Q A L I A 1080
Qy 1074 F T E T V F N Q S T L S A F N G S I P S D K L S T T L E Q A G Y Q O T N Y L P P R T G E D K W V A H G Y D Y G T 1133
Db 1081 F T E T A V L T E S L S A F D G G I T P D E L P G I L T Q A G Y Q O E P V L F P R T G E N K W V A R Q G Y D Y G T 1140
Qy 1134 A A Q W R P O K Q N T O L T G H I T L I W A N Y C V V Q T R A A G L T T S A K Y D W R F L P P V O L T D I N D 1193
Db 1141 E A Q W R P V A Q R N S L L T G R W T L K W D T H Y C V I T Q T O A A G L T V S A N Y D W R F L P T P T Q L T D I N D 1200
Qy 1194 N O H L I T D A L G R P I T L R P W G T E N G W T G Y S S P E K A S P S P D V N A A T E L K K P L P V A Q C V 1253
Db 1201 N V H I L T D A L G R P V T Q R F W G I E S G A T G Y S S E K P S P P N D I D T A I N L T G L P V A Q C L V 1260
Qy 1254 Y A P S W M P L F S Q K T F N R L A E D W K L Y N A R I I T E D G R I C T L A Y R R W Q S Q A I P Q L I S L L 1313
Db 1261 Y A P D S W M P L F S Q E I F N T L T Q E Q E T L R D S R I I T E D W R I C A L T R R R W L Q S Q K I S T P L V K L L 1320
Qy 1314 N G P R L P H S L T L T T D R Y D H D P E Q O I R Q O V V F S D G F G R L L Q A A A R H E A G M A R Q N E D G S L 1373
Db 1321 T N S I G L P P H N I L T A T T D R Y D R S E Q I R Q O V A F S D G F G R L L Q A S V R H E A G E A W Q R N D G S L 1380
Qy 1374 I I N V Q H T E N W A V T G R T E Y D N K G Q P I R T Y Q P Y F L N D W R Y V S D S A R Q E A Y A D T H Y D P 1433
Db 1381 V T K V E N T K R W A V T G R T E Y D N K G Q I R T Y Q P Y F L N D W R Y V S D S A R - - K E A Y A D T H Y D P 1438
Qy 1434 I G R E I K V I T A K G W R F R T L F T P W F T V N E D E N T A A E 1468
Db 1439 I G R E I R V I T A K G W L R Q S Q Y P F W F T V S E D E N T A A D 1473
```

## RESULT 3

US-08-851-567B-32  
; Sequence 32, Application US/08851567B  
; Patent No. 6528484

## GENERAL INFORMATION:

; APPLICANT: Ensign, Jerald C  
; APPLICANT: Bowen, David J  
; APPLICANT: Petell, James  
; APPLICANT: Fatig, Raymond  
; APPLICANT: Schoonover, Sue  
; APPLICANT: Ffrench-Constant, Richard

; APPLICANT: Rocheleau, Thomas A.  
; APPLICANT: Blackburn, Michael B.  
; APPLICANT: Hey, Timothy D.  
; APPLICANT: Merlo, Donald J.  
; APPLICANT: Orr, Gregory L.  
; APPLICANT: Roberts, Jean L.  
; APPLICANT: Strickland, James A.  
; APPLICANT: Guo, Lining  
; APPLICANT: Ciche, Todd A.  
; APPLICANT: Sukhapiinda, Kitisri  
; NUMBER OF SEQUENCES: 88  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Dow Agrosciences Patent Department  
; STREET: 9330 Zionsville Road  
; CITY: Indianapolis  
; STATE: IN  
; COUNTRY: US  
; ZIP: 46268  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/851,567B  
; FILING DATE: 05-MAY-1997  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/063,615  
; FILING DATE: 18-MAY-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/395,497  
; FILING DATE: 28-FEB-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/007,255  
; FILING DATE: 06-NOV-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/608,423  
; FILING DATE: 28-FEB-1996  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/705,484  
; FILING DATE: 28-AUG-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Seay, Nicholas J  
; REGISTRATION NUMBER: 27386  
; REFERENCE/DOCKET NUMBER: 960296.93804  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 608-251-5000  
; TELEFAX: 608-251-9166  
; INFORMATION FOR SEQ ID NO: 32:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1485 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-851-567B-32

Query Match 58.0%; Score 4581.5; DB 2; Length 1485;

Best Local Similarity 57.6%; Pred. No. 0;  
Matches 858; Conservative 224; Mismatches 378; Indels 29; Gaps 11;

Qy 1 M Q N S Q D S I T E L S L P K G G A I T G M E A L T P T P D G M A A L S L P L P I S A G R G V A P A F T L N Y N 60  
Db 1 M Q D S P E V S I T L S L P K G G A I N G M E A L N A A G P D M A S L S L P L S T G R G T A P G L S I Y S 60  
Qy 61 S G A G N S P F G L G W C N V M T I R R T H F G V P H Y D E T D T F L G P E G E V L V A - - - - - D Q P R D E 113  
Db 61 N S A G N G P F G I G W C Q C V M S I S R R T Q H G I P Q Y G N D D T F L S P Q C E V M N I A L D Q G Q P D I R Q D V 120  
Qy 114 S T L Q I N I G A T F T V T G Y R S R L E S H F S P L E Y W Q P K T - - T G K T D F W L I Y S P Q G V H L G K S P 171  
Db 121 K T L Q G V T L P I S Y T V T V T Y Q A R Q I L D F S K I E Y W Q P A S G Q E G R A - F W L I S T P D G H L L G K T A 179

```
172 QY QARISNPQTOTTAQMLLEASVSRGEOIYYQYRAEDDTGCEADEITHHLQATAQRYLHI 231
180 Db QACLAUPONDQQAQMLLEETVTPAGEBHVYQYRAEDEAHCDNEKTAHNVTQAQRYLVQ 239
232 QY VYGNRTASBTLPGDGSAPSQADWLFYLVFDYGERNNLKTTPAFST-TGSMCLCRQDRF 290
240 Db VNYGNIKPQASLFLVDNAPAPPEMLFHLVFDHGERDTSLHTVPTWDAQTAQMSVRPDI 299
291 QY SRYEYGEIRTRRLCRQVLMYHHLQALDLSKITHNGPTLVSRLLINVDISAIASTLVFVR 350
300 Db SRYEYGEVTRRLCOQVLMFHRALTALMAGEASTNDAPELVGRILLEYLDKNAVSVTILIT 359
351 QY RVCHGEDGNVVTLPPELEAYQDESPRHHAHQWPMVDLANENAIQRMQLVDLKGEGPLGL 410
360 Db QLSHESDGRVPTQPLELAWQRDLEKIPTWQBFADLNDNFNSQRYQLVLRGEGPLGL 419
411 QY QDKGAWMYRBAQRLAGEIGSDAVTWKMQPLSVIPSLQSNASLVLDINGDQQLDWITGPG 470
420 Db YDRGAWMYKAPQRBQDGSNAVYDKIAPLPTLPLNLDNASLMDINGDQQLDWVVTASG 479
471 QY LRGHSGRPGSWTRFTPLNALPVEYTHPRAQLADLMGAGLSDLVLIGPKSVRLYANTRD 530
480 Db IRGYHSQQPDGKWTFTFPINALPVEYFHPISIQFADLTGAGLSDLVLIGPKSVRLYANQRN 539
531 QY GFAGKDVVQSGDITLPPVPCADPRKLVAFSDVLGSGOAHLEVSATKVTCPNLRGGRFG 590
540 Db GWRKGEDVPOSTGTLPLVTGTDARKLVAFSDMLGSGQQHLEIKGNKVTCPNLRGGRFG 599
591 QY QPITLPGFSPATFENPAQVYLADLOGSGPTDLYVYHNRDLIFLNKSGNGFAEPTVTRF 650
600 Db QPITLPGFSPENSPNERFLADIDSGTDTLYAQSGLLYLQSGNQFADPLTLAL 659
651 QY PEGRLFDHTCOLQADVQGLGVASLILSVPHMSPHHWRCDLTNNKMWLLNEMNNMGVHH 710
660 Db PEGVQFNTCOLQADVQGLGVASLILSVPHMSPHHWRCDLTNNKMWLLNEMNNMGVHH 719
711 QY TLAYRBSQWLEKAAALTTGQTPVCYLFPFHLTLQWTEDEISGNKLVTLIRVARGA 770
720 Db TLAYRBSQWLEKALQTLKAGSFPACLYFPFHLTLQWTEDEISGNKLVTLIRVARGA 779
771 QY WDRERFRGFGYVEQDTHQLAQGNAPERTTPALTKMYATGLPVIDNALSTEYWR-DD 829
780 Db WDKERFRGFGCIKQTDITTFSHGTAPEQAAPSLSSISWPAFGMDEVDSQLATEYMQADT 839
830 QY QAFAGFSPTTWQ--DNKDVPLTPEDDNRYPNRAKCOLLRASELYGLDDSTNKKVPY 887
840 Db QAYSGFETRYTMDHTNQTQDAFTP-NETQRNMLTRALKGQLRLTELYGLDGTDKQTPY 898
888 QY TVTEFRSQRRLQHTDSRYPVLMSSVVSRYHYERTASDPQSONITLSSDRFGQPLKQ 947
899 Db TVSESRYQVRSIPVNKETELSAMWTALENRSYHYERTITDPQFSQSLKQHDIFGQSLQS 958
948 QY LSVQYPRQOPAINLYPDTLPDKLANSYDDQORQLRLTYQSSWHHLTNNTVRVLGLPD 1007
959 Db VTAMPREPRANVPPTLPETLFDSSYDDQQLLALVRQKSWHLTDGENWRGLPN 1018
1008 QY STRSDIFTYGAENVAPAGLNLLELSDKNSIADDPREYLGQQTATYDQNTPLQPT 1067
1019 Db AQRDVVYDRSKIPTEGISLEILLKDDGLADEKAAVYLGQQQTFYTAGQAEVLEKPT 1078
1068 QY RQALIAETEITVFNQSTLAPNGSIPSDKLSLTLBOAGYQCTNYLFPRTGEDKVMVAHG 1127
1079 Db LQALVAFQETAMMDTSLQAYEGVIEBOELNTALTQAGYQOQVAFNFRSPVMAARQ 1138
1128 QY YTDYGTAAQFWRPQKQNTQLTGKITLIWDANYCVVYQVTRDAAGLTTSKTYDMRFLFPVQ 1187
1139 Db YTDYGTAAQFWRPQKQNTQLTGKITLIWDANYCVVYQVTRDAAGLTTSKTYDMRFLFPVQ 1198
1188 QY LTDINDNQHLITLDALGRPITLRFWGTENGQHTQYSSPEKASFPSPSDVNAIEBLKPLP 1247
1199 Db LTDINDNQHLITLDALGRVITLRFWGTENGQHTQYSSPEKASFPSPSDVNAIEBLKPLP 1255
```

---

```
1248 QY VAQCQVYAPESWMPVLSQKTFNRL---AEQWQKLYNARIITEDGRICTLAYRRWVQSOK 1304
1256 Db VAQCLVYVDSWMPVLSLSQLSQSEAEALWAQRAAHMITEDGKVCALSGKRGTSQHN 1315
1305 QY AIPQLISLNNGRPLPPHSLTLITTDYHDHPQOIRQOVVFSDFGFRLLLOAAARHAGMA 1364
1316 Db LATQLISLNASIPRLPPHVLGITTDYSDSDPQQHQOQTVSFSDFGFRLLQSSARHESGDA 1375
1365 QY RORNEGSLIINVQ-----HTENRWAVTGRTEYDNKGOPITRYQPYFLNDRMYVNSDA 1418
1376 Db WORKEGGLVVDANGVLSVAPDITRWAVSGRTEYDDGQPVRTYQPYFLNDRMYVNSDA 1435
1419 QY RQKEAYADHYVDPITGREIKVITAKGWRFRRLFTPMFTVNSDENTAA 1467
1436 Db RQKEAYADHYVDPITGREIKVITAKGWRFRRLFTPMFTVNSDENTAA 1482
```

RESULT 4  
US-09-492-709A-340  
; Sequence 340, Application US/09492709A  
; Patent No. 6720139  
; GENERAL INFORMATION:  
; APPLICANT: ZyeKind, Judith  
; APPLICANT: Ohlsen, Kari L.  
; APPLICANT: Trawick, John  
; APPLICANT: Forsyth, R. Allyn  
; APPLICANT: Froelich, Jamie M.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: GENES IDENTIFIED AS REQUIRED FOR PROLIFERATION IN  
; FILE REFERENCE: ELITRA.001A  
; CURRENT APPLICATION NUMBER: US/09/492,709A  
; CURRENT FILING DATE: 2000-01-27  
; NUMBER OF SEQ ID NOS: 485  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 340  
; TYPE: PRT  
; LENGTH: 1426  
; ORGANISM: E. Coli  
US-09-492-709A-340

Query Match 2.0%; Score 160.5; DB 2; Length 1426;  
Best Local Similarity 19.0%; Pred. No. 0.00045;  
Matches 255; Conservative 152; Mismatches 446; Indels 487; Gaps 65;

```
406 QY LPGLLYQDKGAWMYRBAQRLAGEIGSDAVTWKMQPLSVIPSLQSNASLVLDINGDQQLDW 465
70 Db LPFILSRITYSSYRTKTPAPVGVFGPG---WK--APSDIRLQLRDDGLILNDNGRSIHFE 124
466 QY ITQPG-----LRGHSQRPDGSWTRFTPLNALPVEYTHPRAQLADLMGAGLS 513
125 Db PLLPGEAVYSRSESMMLVRGKAAQPDGH-----TLARLWGLALPPD 165
514 QY LVLLIGPKSVRLYANTRDGRFAGK-----KDVQSGDITLPPVGDAPRKLVAFSDVLG-- 564
166 Db IRL-----SPHLYLATNS--AQGPWWILGWSRVPGAEDVLPALPPYRVLTGHWADRFRT 219
565 QY -----SQGAHLEVEVSATKVTCPNLRGGRFGQPT 594
220 Db LTYRREAAGDLAGEITGVTGAGREFRLVLTQQAAREBARTS-----SLSSDSSRPLS 274
595 QY LPQF--SQPATEPN-----AOVYLADLQSGPTDI-----IYVHNRDLIFLNK 638
275 Db ASAFPDTLPCTGYGPGRGIRLSAVLMLH-DPAYPESLPAPLVRYTYTEAGELLAVYDRS 333
639 QY GNGFAEPTVTLRFPEGLRFDHTCQLQADVQGLGVASLILSVPHMSPHHWRCDLTNNKMW 698
334 Db NTQV-----RAFTYD-----AQHPRVVAHYRAGRPEM---RYDYDTGR---- 370
699 QY LNEMNNMGVHHITLRYRSSSQWLEKAAALTTGQTPVCYLFPFHLTLQWTEDEISGN 758
```



371 VVEQLNAGLSYRYLYEQ-----DRITVTDLSNRREV-----LHT-----EGGAGL 411  
Qy 759 KLVTTLYARGANDGRERERFGRFGVGTQDSHQIQAQNAQNPARTPPALTKWYATGLPVID 818  
Db 412 KRUVVKELADGS-----VTRSGYDAAGRLTAQTAAGRTTEY--GLNVVS 454  
Qy 819 NALS-----TEYWRDD--QAFAGSPRFTWQDNKQVPLTPEDDNRWYFNWRAKLG 867  
Db 455 GDITDIITPPGRETFTYNDGNQUTAVVSPDGLSRREYDEPGRLVSETSR-----SG 507  
Qy 868 QLLASLEYGLDDSTNKHVPYVTVTFBRSQVRRLQHTDSRYPLVMSVVSRESRNYHYERIASD 927  
Db 508 ETVA--YRYDDA-HSELPAITTDATGSTRQ-----TWS-----RYGQLLAF 546  
Qy 928 PQCSQNT-LSSDFGQPLKQLSVQYPRQOPAINLPDPLDKLLANSYDDQORQLRLT 986  
Db 547 TDCSGYQTRYEYDRFGQ-----MTAVHREBEGISLYRRYDNRGLTSVKDAQRETRYE 599  
Qy 987 Y-----QQSSWHHLTNTTV-----RVLGLPDSTRS-DI 1013  
Db 600 YNAAGDLTAVITPPGNRSETQYDAGKAVSTTQGLTRFSMEYDAAGRVISLTNENGSHSV 659  
Qy 1014 FTYGAEN--VPAGLN-----LELLSDKNSLIA-----RVGQLLAF 546  
Db 660 PSYDALORLVQGGFGDGTQRYHYDLTGKLTQSEDEGLVILWYDESDRITHRTVNGEPA 719  
Qy 1040 -----DQKPR-----EYLGQOKT--AY 1054  
Db 720 EQWQYDGHGMLTDIHLSEGRVAVHYGYDDKGLTGECQTVENPETGELLWQHTKHAY 779  
Qy 1055 TD---GQNTTLPQPTPQ-----ALIAFTETVFNQSTLS--AFNGSI 1092  
Db 780 NEOGLANRVTPDSLPFVWELTYGSGYLAKMGLGTPLVETRDLRHRETVRFSFGMAGSN 839  
Qy 1093 PSOKLSTTLTGAGYQQTNYLPPRTGEDKQVVAHGYTDYGTAAQFWRPQ--KQNTQLTG 1150  
Db 840 AAYELTSTYTPAGQLQSHL-----NSLVYDRDYGSDNGDLVRISGPRQTRYGYGATG 894  
Qy 1151 KI---TLWDANYCVVQVTRDAAG-----LTSKAYDMRFLTP 1185  
Db 895 RLESVRTLPDLOIRIPVAT--DPAGNRLPOPELHPDSTLTWPDPNRAEDAHYVYRDEY 953  
Qy 1186 VQLTIDN-----DNQHLITL-----DALGRPIT 1208  
Db 954 GRUTEKTRDIPAGVIRTDERTHYHDSQRLVFTYRIQHGEPLVESRYLYDPLGRMA 1013  
Qy 1209 LRFWGTENGKMTGYSS---PEKASFSPSPDVNAALIELKXPLPVAOCQ-VYAPESWMPVL 1263  
Db 1014 KRWWRER-DLTGHWLSLRKPEVTWYWGDDGRLTVQ-----TDTRIQTVEPGSFTPLI 1068  
Qy 1264 SOKTFNRLAEQDMQKLYNARIITEDGRICTLAYRRWVQSOKAIPQLISLNNNGRPLP-PH 1322  
Db 1069 RVETENGEREKA-QRRSLAETLQEGS-----ENGHVVFPA 1104  
Qy 1323 SLTLTTRYDHPDQOQTRQVQVDFSDGGRLLQAAARHAGHAGARQNE--GSLIINVQHT 1380  
Db 1105 ELVRLDLRL---EEIRADRVSSESRAWLAQCGLTVEQ-LARQVPEYTPARKAHLYHC 1159  
Qy 1381 ENR-----WAVTGRTEYD-----NKGOPIRTYQPYFELNDWRYVNSDSARQ 1420  
Db 1160 DHRGLPLALISEDGNATWS---AEDWGNQNLNEENPHVYQYRPLPGQOH-----DE 1209  
Qy 1421 EKEAYADTH-VYDPI-GREI 1438  
Db 1210 ESGLYNRRHYDPLQGYI 1229

## RESULT 5

US-09-252-991A-24910  
; Sequence 24910, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252.991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 24910  
; LENGTH: 658  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
; US-09-252-991A-24910

Query Match 2.0%; Score 155.5; DB 2; Length 658;  
Best Local Similarity 20.1%; Pred. No. 0.00031;  
Matches 113; Conservative 61; Mismatches 155; Indels 233; Gaps 31;  
Qy 356 QDGNVTVLPLELAYQDFSPRHHAHQPMQMDVLANFNAIQWQLVDLKGGL-PGLLYQ-- 412  
Db 123 REGDNVTLVRNRRLAEDTS---IHWGIIILPANMDGVP---GLSFEGIAPGGLYEYR 173  
Qy 413 ---DKGAWWYRQAORL-----GEIGSDA-----VTHEKQPLSVIP 445  
Db 174 FKYRONGTYWYHSHGSLQEQAGVYGALVIDAREPEPFSYDRDYVYVLLSDMSDEKQRI 233  
Qy 446 SLOSNAS-----LVDINGCQOLDWVIT-----GP-----GURGY-- 474  
Db 234 KLKQSDYNNFHKRTVCGDFIDVVSANG---WAATLADRKWAEMKMSPTDLADVSGTYT 290  
Qy 475 ---HSQRPDGSGWT-----RFTPLNAL-----PVEYT 497  
Db 291 YLLNGQPPDGNWTGLPRGPKLRFRVNASMSYFVRI-PGLKMTVVAADQHQHVEPVSD 350  
Qy 498 HPAQALADLACAGLSLDLVILGPKSVRLY-----ANTRDGAQKGVQVQSGDITLIPVQAD 552  
Db 351 ELARIAVAETVD-----VIVEPGRRAYTLFAQSMDSRGVARGTLALAEGLSAPVPTPD 403  
Qy 553 PRKLVAFSDVLGSGQAH-----LVEVSATKVTCPNLG----- 585  
Db 404 PRPLIGMDDMGCMGDHGAHGAAATPASEMDHSRSMGMDKMGMDHSHKAGMDMNGMDH 463  
Qy 586 ---RGRFGQPIITLPGFSQPATE-FNPAQVYLADLGGSGPTDLIYVHTNRLDIFLNK 637  
Db 464 SKMAGMDHSMGMG-AMPMQSHPASEDGNP---LVDMQTWTPTPKL-----ADPGLGL 512  
Qy 638 SGNG-----FAPVTVLRF--PGLRFDHTCQLQADVQGLGVASLILSVPHMSPHHWRCD 690  
Db 513 RDNGRRVLYTAD-LRSRFADPDGREGPTIELHLTG-----HMEKFANFSFD 557  
Qy 691 ---LTNWKPLWANE-----MNNMNGVHTLRYRSSSQFWDLDEKAAALTTGQTPVCYL 739  
Db 558 GIKFSDAEPRLTYGERLRLVNDTMTHT----- 587  
Qy 740 PFPHTLMQTEDEISGNKLV 761  
Db 588 PIHLGMMW-SLDEDE-QGNEFLV 607

## RESULT 6

US-09-543-681A-5434  
; Sequence 5434, Application US/09543681A  
; Patent No. 6605709  
; GENERAL INFORMATION:  
; APPLICANT: GARY BRETON  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS  
; FILE REFERENCE: 2709.1002-001  
; CURRENT APPLICATION NUMBER: US/09/543.681A  
; CURRENT FILING DATE: 2000-04-05  
; PRIOR APPLICATION NUMBER: US 60/128,706

; PRIOR FILING DATE: 1999-04-09  
 ; NUMBER OF SEQ ID NOS: 8344  
 ; SEQ ID NO 5434  
 ; LENGTH: 2315  
 ; TYPE: PRT  
 ; ORGANISM: Proteus mirabilis  
 US-09-543-681A-5434

Query Match 2.0%; Score 155.5; DB 2; Length 2315;  
 Best Local Similarity 20.0%; Pred. No. 0.003;  
 Matches 313; Conservative 179; Mismatches 558; Indels 513; Gaps 81;

Qy	10	TELSLPKGGGAITGMEALNFTGPD-----GMAALSPLP-PISAGRY 51
Db	662	TEVTVP--GGAAGDTLTLITIPDGSOTVEHTLTADEVTAAGKADVTIPADKVTADGNY 719
Qy	52	APAFLLNYSNGAGNSPFLGMDNCNMVIRRTIRHFGVPHYDETDFLGEGEVLVADQPR 111
Db	720	--SVTAETDPAGNTS--GGQ-----KPTDFAVDTVAPSAPVLKAEDDGSVVDLPT 767
Qy	112	DESTLOGINLGATFTTGRSRLSHESRLEYPQKTTGKTFWLIYSPD-----161
Db	768	D-----ANKGDTVEITFEDEKDKHTVTLKGDNGMTSDTPALI---PDSNGDKATIPA 818
Qy	162	-----GOVHLGKSPQARISNPSTQTOTAQMLLEASVSRGEQIYYQYRAEDDTGCEAD 215
Db	819	DNVKDSEVTGIAKDPGNSDESFTVTSKTDGVADAPVLTPEVA-----DGVANAD 870
Qy	216	EITHLQA-----TAQRYLHIV-----232
Db	871	ELKDGLOAEVTLPGTVEGAETLTVTRPKTTETVTHVTKDAAAGKVSVDIPKDAVQ 930
Qy	233	-----YYGNRTASETLPG-----LDGSAQADMLFVLVPDYGRSNNLTPPA 276
Db	931	NGQNSVDVSLTQGNPAK---PGNKVDFAVDGQIPGDTD-----GDGTVDTPV 976
Qy	277	FSTTGMCLQRDRFSRYEYGFETRLCRQLVMYHILQALDS---KITEHNGPTLVSR 333
Db	977	VTIPEA---ADGVNADELKGVQT-----QVTVPGSAGDTLTLITIPDGSOTDVEH 1027
Qy	334	ILNYDESAIATLVFVRVGHQDGNV-----VTLP-----PLELAYQDFSPRHH 379
Db	1028	TLTADEVTAAGKADVTIPADKVTADGNSVTAEITDPAGNTSGQKPTDFAVDTVAPS---1084
Qy	380	HWQP-----MDVLANFAIQRLV---DLKGEGLPGLLYQDKGAWTWSAQLRG 426
Db	1085	--APVLKAEDDGSVVDLPTDANKGDTVEITFEDEKDKHTVTLKGDNGMTSDTPALIP 1142
Qy	427	EIGSDAVT--WEKMOPLSVPLS-----QSNASLVINGDQDLM--VITGQLR--GYH 475
Db	1143	DSNGDKATIPADNVKONSEVTGIAKDPGNSDESFTVTSKTDGVADAPVLTPEVADGYA 1202
Qy	476	S--QRPDGSTRFTPLNALPVEYTHPRAQLADMLGAGLSDLVLIGPKSVRLYAN--TRDG 531
Db	1203	NADBLKDG-----LQAEVTLPGATIE---GABITLTVTRPKTTETVTHVTKDE 1249
Qy	532	PAKG-----KDVQSG-----DITLVPVQADPRK---LVAPS-----DVLGSGAHL 570
Db	1250	AAAGKVSVDIPKAVQNGQNSVDVSL--TQGNPAKPGNKVDFAVDGQIPGDTDGDG---1304
Qy	571	VEYSATKVTCPNPLRG-----RFGQPTLTPGFSQAPATEFNPAPVYLAIDLGSQPTD 622
Db	1305	-TVDTTPVTIPEATDGVNADELKGVQTEVTVPVGGSAAG---DTLTLITKPGDS--TD 1358
Qy	623	LI-----YVTHNRLDIFL-----NKSQN-----GPAEPVTLRFPGLRPHDTCQLQM- 664
Db	1359	TVHEHTLTADEVTAAGKADVTIPADKVTADGNSVTAEITDPAGNTSGQKPTDFAVDTQIP 1418
Qy	665	ADVQGLGVASL--ILSVPHSPHWRCDLTNPKFWLLNEMNNMNVGHHTLYRSSSQFWL 722
Db	1419	GDTDGDGVDPVTPVTIPEAT-----DGVN-----ADELKDGVQTEVTVPVGGSA---1463
Qy	723	DEKAAALTGQTPVCYLPFP---IHTLWQTEDEISGNKLVTLTRVARGAWDGREREFR 779

Db	1464	-----GDTLTLITKPDGSDTVEHTLTADEVTAAGKADVTIPADKATPDG-----1508
Qy	780	FGGYVEQTDHQLAQGNAPERTPPALTKNWYATGLPVI---DN-ALSTFYWRDDQAFAGF 835
Db	1509	--NYSVKAETDPAGNTSGEKATDFTVDTAPSTPVLNAEDNGSVSVLPDGD--ANKGD 1564
Qy	836	SPRFTTWQDNK-----DVLTPEDDNRWYFNRAKQQLLR--SELY 875
Db	1565	TVETI-TFEDEKDKQTVTMEKGDNGWTSSDPNLIIPDSQN-----NTAISDNVKNSEVT 1619
Qy	876	GL--DDSTNKHVPYVTFEPRSQRRLQHTDSRYPLVMSVSVSRNHYHRIASDPCCSQN 933
Db	1620	ATAKOPSGNESAPATA-----MSKTDVLPVTSISVDTTSVNDNGDKISADASVSGE 1670
Qy	934	ITL-----SSDRFQQLKQLSVQYPRRQQAIIYDPTL-----PD-KLLANSY 976
Db	1671	ITEVPATIBDKDDTTGL-VYTVVALDYAAQDVTVTI---TLTNDAGHASAPDYSTLAGSQ 1726
Qy	977	DDQOROL-----RLTYQOSSWHHL-----995
Db	1727	HGKIALHGDGKVTYDGAFTVTVIPAGSKSVSFIVDPTMEANQNAFNAEGKVVATI 1786
Qy	996	-----TNNTVRVL-----GLPDSTRS-----DIPTYGAEN-----1020
Db	1787	TGTSNNATAVTDVNNAGASATGVIYDGNPISLRNLDGDTLKYSLSSVAENGDFGYTI 1846
Qy	1021	VPAGLNLLELSDKNSLIADDPREYL-----GQKQYAT-----DQONTTLPQ 1064
Db	1847	GVDGSKDKPLTVDYNDTI-----YVGYYQDGSSTSSYNSLANSQDNGPDGKTGDNQ 1899
Qy	1065	TPTRQALIAFTETTFVNQSTLS-----AFNGSIPSDKSLSTTLEQ--AGYQQTNYLFP 1115
Db	1900	SITVDLGGAGDLMVIRGNLNTNRYVAGEGNDTFTMDGNTALRSWYAG---SYLFME 1955
Qy	1116	TGEBKVMVAHHGYTDYQ-----TAAQFMRPQKSNLTQLTGKITLWDANYCVVVQT 1166
Db	1956	SGNDVTIKRTGVTNAGQIVLGSQSDTFQ--GDANDQNNTELGLDL-----2002
Qy	1167	RDAGLTTSKDYDRFLTPVQ-----LTDINDNQHLITLALGPITLRFWGTENGK-MTG 1221
Db	2003	--GSGTQDKSNMPEYLYQDGSNLGNDN---NIDAVTDVNTVTIYGSVGBILGG 2056
Qy	1222	YSS 1224
Db	2057	YGS 2059

RESULT 7

US-09-091-609-2  
 ; Sequence 2, Application US/09091609  
 ; Patent No. 6600029  
 ; GENERAL INFORMATION:  
 ; APPLICANT: SHERMAN, DAVID H.  
 ; APPLICANT: WILLIAMS, MARK D.  
 ; APPLICANT: XUE, YONGQUAN  
 ; TITLE OF INVENTION: METABOLIC ENGINEERING OF  
 ; FILE OF INVENTION: POLYHYDROXYALKANOATE MONOMER SYNTHASES  
 ; FILE REFERENCE: 600.297US2  
 ; CURRENT FILING DATE: 1998-06-19  
 ; EARLIER FILING DATE: 1996-12-18  
 ; EARLIER FILING DATE: 1996-12-18  
 ; EARLIER FILING DATE: 1995-12-19  
 ; NUMBER OF SEQ ID NOS: 4  
 ; SOFTWARE: FastSeq for Windows Version 3.0  
 ; SEQ ID NO 2  
 ; LENGTH: 4630  
 ; TYPE: PRT  
 ; ORGANISM: Streptomyces venezuelae  
 US-09-091-609-2



```
QY 661 QLMADVQGLGVASLILSVPHMSPHW--RCDLTNMK-----PWLLENMNNMVGHTLRY 714
DB 2442 VLVGTGGGLGA-----LLARHLVERHCVRLLLSVRGADAPGAADLGEDLAGAEVAF 2497
QY 715 RSSQFWLDBKAALTT--GQTPVCVLPFPIHLMOTE--TEDEISGNKLVTLR--YARG 769
DB 2498 AAADAADRESLARAIATVPAEHP---LTAVVHTAGVVDDATVEALTPERLDAVLRPKVDA 2554
QY 770 AWDGRE--REFRGFGVYQSDSHQL-----AQGN--APERTPPALTKWYATGLPVIDNAL 821
DB 2555 AWWLHETKDLRLDAFVFSFSGVGTAGQAAANTGLDAAHRAATGAA--TSL 2612
QY 822 STEYWRDQAFAG--SPRETTQDNKDVLTPEDDNSRYWFNRALKQGLRSELYGLOD 879
DB 2613 AWGLNDGTHGCGTGLGAADLARSRAGITPLP-----LQGLALPDAAVARD 2660
QY 880 STNKHVPYVTFERSQVRLRQH--TDSRYPLMSSVVESRNYHYERIASDQCSQNTLS 937
DB 2661 AL--LVB-----AGLRPTAHRGTDGQPPALWRGLVRARPRAARTAAE-----A 2702
QY 938 SDRFGQPLKOLSVOYP--RRQOPAINLYPDTLPDKLLANSYDDQORQLRLTYQSSWHHLT 996
DB 2703 ADTTGGWLSGLAAQSPERRSTAVTLTVGVAD----- 2735
QY 997 NNTVRVLGLPDSRSDIFTYGAENVAGGLNLELLSKNSLIADDKPREVLGQOKTAYTD 1056
DB 2736 -----VLGHADSA-----AVGAER-----SFKD-----LGFDSLAGE 2763
QY 1057 GQN-----TTPLOTPTRQALIAFTETTVFNOSTLSAFN-----GSIPIKLSL--TLEQAGY 1106
DB 2764 LRNLNAATGLRLPA-----TVFDPHPSAALASHLLAQVPGLEKGTAAATVVAE 2814
QY 1107 QQTNYLFPRTGEDKWVAHGYDYGTAQF-----WRPQKQSNLTG-KITLI 1155
DB 2815 RGASFQDRATDDDFIAIV-----GMACRYPGGVSSPDLWRLVAREGTDIAISEFPVNRG 2867
QY 1156 WD-----ANYCVVQTRDAAGLTSAKYDWRF--LTPVQLTIDINDQHL-- 1197
DB 2868 WDLSELYDPDPESKGYTYC-----REGGLEAGDFFAAPFGISPREALVMDPQORLLE 2922
QY 1198 ITLDALGR-----PITLRFWGTENGKMTGYSSPEKAS 1229
DB 2923 VSWEALERAGIDPSSLA--GSRGGVYVGAAGHSYAS 2956

RESULT 9
US-09-543-681A-7181
; Sequence 7181, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 7181
; LENGTH: 1028
; TYPE: PRT
; ORGANISM: Proteus mirabilis
US-09-543-681A-7181

Query Match 1.9%, Score 153.5; DB 2; Length 1028;
Best Local Similarity 18.0%; Pred. No. 0.001;
Matches 204; Conservative 124; Mismatches 356; Indels 449; Gaps 52;

QY 19 GAITGMEALPTGPDGMAAL-----SULPLISAGRGVAPATLNNYGAGNSPFGLG 71
DB 162 GSATGKGLDITAPDQAGTLFFDQIIMSVPD---NRWFIPDYQIPYVNNAVNTMYSKN 218
```

```
QY 72 WDCNVMITRRRTHTFGVPHYDETDITFLGPEGEVLVWADQPRDESTLQGINLGATFTVTGYR 131
DB 219 WSALLM-----YDQMFTAHP---TLNFATEFRDD-----Q 246
QY 132 SRLSHFSRLVWQPKTKTKTDFWLIYSPDQVHLGKSPQARISNPSTOTOTAQMLLEA 191
DB 247 PEVASIYQRFYYQ-----GVSRRDKKITAEIMDKNLALM----- 280
QY 192 SVSSRGEOIYYQVRAEDDTGCEADEITHHLQATAQRYLHIVVYGNRTASSETLPGLDGAP 251
DB 281 -----KKLALQSHA----- 294
QY 252 SOADMFLYLVFOYGERSNMLKTPAPSTTGSWMLCRODRFSRYEYGEPIRRLRCROVLAY 311
DB 295 AXA-----LDHPNQNFIKVEGVSE-----ETQK----- 319
QY 312 HHLQALDSKITEHNGTLV-----SRLLNVDSEIASITLV-----FVRVGH 354
DB 320 ---ALLDANMLRDVGKTLQTAIYLRSRSLSAIDRKLETLVLLGTRYVLEQGFTRGSGY 376
QY 355 EQDGNVVTLPPLLEYAD-----PSPRHHAHQMPMDVLANFNATQRMQLVDLKGELP 407
DB 377 Q-----IIT-----HVGQYTRLEFDAMFIGRH-----ILAKHNLAPLPTQ 410
QY 408 GLLYQDKGAWYRSARQLGEIGSDAVTWBKMOPLSVIPSLQSNASLVLDINGD---QQLDW 464
DB 411 -----QAMWYNATGRI-----FEKDNEIVDANVDILNTQLQW 443
QY 465 VITGPGLRGYHGORPDG---SWTRFTPLNALPV-----EYTHPPRAQLADLM 507
DB 444 MIKSLMLPDYQORQQAQLAQSLWLNKLTILSSKGAVAGFKSDGISIFPHSHQHPAYAKDAF 503
QY 508 GAGSLDLVLIGPKSVRL---YANTPDGPAKGVQVQSGDITLPVPG-----ADP 553
DB 504 GGLAPSVYALSHSPFLSSPAHARLKVLLKRIYTKETQIPLVLSGRHPTGLHKISIDP 563
QY 554 RKLVAFSVDLGSQAHLVEVSATKVTCPNLGR-----GRFGQPIITLPGFS 599
DB 564 FKWMALAGTPDGQK---ELDTTLAAAYAKLANKDSFEGIKAENEPVGAWMYASMAIQ 619
QY 600 QPATEFNPAQVYLADLDGSGPTDLI-----YVHTNRLDIF---LNKSG----- 639
DB 620 RRASITAPQOSWLAIAARGFS--RYLVGNESYENNNRYGRYLQYQGLRILIPADLTGSGFSA 678
QY 640 ---NGPAEPVTLRFPBGLRPFHTCQLQMAVQGLGVASLILSVPHMSPHHWCIDLTMK 695
DB 679 GNDWNPYGTTHILP---YDE--LEAKUSQLPSAGIEMLLSTQRYSG----- 722
QY 696 PWLLNEMNNN---MGVHHTLRYR-----SSSQFWLDBKAAALTTG-----QTPVC 737
DB 723 ---ANTLNNSMFAKMLGHGSKYQQQSLRANKSYFLFDNRVIALGSGIENNDKQHTTET 779
QY 738 YLPFPPIHTMQTETEDEISGNKLVTLRLYARGAWDGRERFRGFGVYVQPDHQLAOGNA 797
DB 780 LFOFAVFKL-----QSIILNGKKVNL-----GTQLTLNNA 810
QY 798 PERTPALTKWYATGLPVIDNALSTEYWRDD-----QAFA-----CFSPRFTWQD 844
DB 811 DTLIDPAGNLYKLAAG--QTVFESYQKQYSDVDRNSQOTEQLFATAVISHGKAPK----- 863
QY 845 NKDVPLTPEDDNSRYWNRALKQGLRSELYGLDDSTNKHVPYTVTFEFSQVRLQHTDS 904
DB 864 -----NANYEYATAIEAQ-----DNKAPETV-----LQHNQ 891
QY 905 RYPLMSSVVESRNYHYERIASDQCSQNTILSSDRFGQPLKOLSVOYPRRQ-----QP 958
DB 892 LHAVK-DKITQEBGYAFEN-ATEVNSSQALLSSD---SP-TMWVVKQKQQLTSLVNP 945
QY 959 AINLYPDTLPDKLLANSYDDQORQLRLTYQSSWHHLT-----NNTVRVLGL 1005
DB 946 DLNLYOGIEADQI-----DNKGQVEVSYSROW---LTADPPQISSITVTVKGI 991
```

```

RES021.10
US-09-711-164-467
; Sequence 467, Application US/09711164
; Patent NO. 6589738
; GENERAL INFORMATION:
; APPLICANT: Forsyth, R. Alllyn
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; TITLE OF INVENTION: GENES ESSENTIAL FOR MICROBIAL PROLIFERATION AND ANTISENSE THERAPY
; FILE REFERENCE: ELITRA 008A
; CURRENT APPLICATION NUMBER: US/09/711.164
; CURRENT FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: US 60/164415
; PRIOR FILING DATE: 1999-11-9
; NUMBER OF SEQ ID NOS: 469
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 467
; LENGTH: 1377
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-711-164-467

```

Query Match 1.8%; Score 144; DB 2; Length 1377;  
Best Local Similarity 18.6%; Pred. No. 0.012;  
Matches 230; Conservative 133; Mismatches 403; Indels 472; Gaps 62;

406	Qy	LPGLLYQDKGAWWTRSAQRLEIGSDAVTWKQMPPLSVIPSLQSNASLVINDINGDQGLDWV	465
70	Db	LPFTLSRTYSYRTKTPAPVGSGLPG--WK--MPADIRLQLRDNTLILSDNGSRSLYPE	124
466	Qy	ITGPCLRGYHSQRDPGSGWT-----RFTPL-NALPVEY-THPRAQLAD-----	505
125	Db	HLFPGEDGY--SRSESLWLVRGSGVAKLDEGHRLAALWQALPEELURLSPHYRLATNSPQGP	182
506	Qy	--LMG-----AGLSPLDLVILGP-KSVRLYANTRDGFAKGKDVVO-----SGDITLPPVPGAD	552
183	Db	WLLGWLCERPEADEVLPALPPYRVLTGLVDRFRGTQTHREAAAGFSGSEITGVTDGAW	242
553	Qy	PR-KLVAFSDVLGSGQAHLEVS--ATKWTCPNIGRGRFGQPITLPGFSOPATEFNPQAV	610
243	Db	RHFLVLTTQAQRAEBAARQAISGGTEPSAFFD-----TLPGVYTB-----	282
611	Qy	YLADLGGSGPTDLIYVHTN--RLD-IFLNSKNGFAPVTLRFPFGL-----RFDHTQC	661
283	Db	-----YGRDNGIRLSAVWLTH-----DP--EYENLPAAPLVRYGWTPTP	319
662	Qy	LQMA--DVOGLGVASLILSVPH--NSPH-----HWRCDLTNKKPWLLEMMNNMG	707
320	Db	GELAVYDRSGKQVRSYTDYDKYGRGVMAHHTGRPEIRYRYSDBG-----VTQLNPAG	375
708	Qy	VHHTLRY---RSSQFQLDEKAAALTTGQTPVCYLPPLPPIHTLWQTEDEBSGNKLVTTL	764
376	Db	LSYTYQYEKDIRITITDSLDRREVLTQGE-----AGLKRVPVKK	413
765	Qy	RYARGAWDGREREPFRGFGYVE-QTDSHQLAQGNAPERTPPALTQKWYATGLPVIDNALST	823
414	Db	EHADGS--VTQSQPDVAGVLLRAQTD---AAGRITTEYSPDVVT--GLITRITTPDGRASA	465
824	Qy	EYWRDDQAAPAGSPRPTTWQDNKV-----PLTPEDDNSRYFWNRALKGQLLR	871
466	Db	FYNNHHNQLTS-----ATGPDGLREIREYDELGRLIQETAPDGIITRY-----	510
872	Qy	SELYGLDDSTNHKHPVYVTEPRSQVRRLQHTDSRYPVLWSVSVESRNVHYERIASDPQCS	931
511	Db	-----DNPHSDUP-CATE-----DATGSRKWTWS-----RYGQLLSFTDCS	546
932	Qy	QNIT-LGSDRFGQ-----PLQLSVQYPRRQQPAINLYPDLPKLLANSVDDQQRQ	982
547	Db	GYVTRYDHDREGMATVHREBGLSQRAYDSRGQLIAVK-----DTQGE	591
983	Qy	LRLTYQOOSWHHLTNNTVRVLGLPDPSTRSIDIFTFYCAENVPAAGJUNLELLSDKNSLIADKK	1042
592	Db	TRYBY-----NIAGDLTAVIAPDG-----	610

## RESIT.T 11

```

RESULT 11
US-09-489-039A-10045
; Sequence 10045, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSIS AND THERAPEUTICS
; FILE REFERENCE: 2709, 2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1993-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 10045
; LENGTH: 798
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-10045

```

```

Query Match      1.8%; Score 143; DB 2; Length 798;
Best Local Similarity 20.9%; Pred. No. 0.0056;
Matches 137; Conservative 81; Mismatches 232; Indels 206; Gaps 34;

Qy      18  GGAITGGEALTTGPDGMAALSLPISAGRCYAPAPTLNYSAGNSGSPGLGWDC--- 74
Db      123  GYNIRGMESNRVGDVDGIAQ-----PNATGRGV-----GRAGLNTFGIGRDYIDP 169

Qy      75  -----NVMTIRRRTHFGV-PHYDET----- 94
Db      170  YMTGSDVIQSGATSTETANSAIGNVSFPRKSADDYLRPKTSAFGYRSGYDSADRSWHN 229

Qy      95  --TFLGPEGEVL--VVADQPDRDESTLQGINLGATFTVTGYSRLESH--FSRLEYWQP--- 146
Db      230  GVTAVAGD--EFLRGILVYSRSDGOETN--NSG---TVDAVPANWHSDAFLASGIWQPNDE 284

```

147 -KTTGKTDWLIYSPDQVH--LLGKSPQARISNPSTOTTAQWLLAS-----VS 194  
285 HKLSTSPDY---YKTNHHTYDWDSSGNSGTIGTANQTSQTRWGLSKDDWTMPMDYLD 341  
195 SRGEQYYQVRAEDDTCGEADETHHLQAQRYLHIVYGNRTA-SETPLGLDGSAPSQ 253  
342 SVSTKIYYQTEAHDWTYMPDSVTRXMQTVNSNY-DTDTWGLQTLAKTLGRHDLA--- 397  
254 ADWLFLVLPVDPYGRSNNLTPPASTTGSWLCQDRFSRYEFTETRLCLCRQVLMYHH 313  
398 -----GNASTSKTQRPFS----- 411  
314 LQALDSKITEHNGFTVLSRLNLYNDESAIATLTVFVR-RVGHEODG-NVVTLPPELAYQ 371  
412 -----QSPI-----PSVYSEIMQEADESRSTGLGFQVDKINPDLDSHFVIFGVRVHQ 462  
372 DFGPRHHAHQMDVLNFAINQWOLVDLKGEG-----LPGLLYQ-----D 413  
463 STKPENLS-----DLAANSVLSESSVANLYGKNSDTQVLPSTFQYDLTPRLMTYLYQY 517  
414 KGAWVRSARQLG--EIGSDAVTWKMOPLSVIPSLQSNASLVINDGDCGLDHWITGPGL 471  
518 RGAQFPNASQLYGSWNLGS---SYAGSQYALI-----GNTDLKTETSD-NLEWGLKGEVT 569  
472 RG-----YHSORPDGWSRFTPLNALPVEVTHPRAQLADLWAGLSDLVLI--GPKS 521  
570 EGIETRLTALFYNYSKQFIATRTTRANN-PQGTINVPNSNTIYQAEHRDKAVIYGEIS 628  
522 VR-----LYANTRDGFAGKGVVQ--SGDITLPPVPGADPRKL-----VAFSD 561  
629 TKFNFGTFWQVQDLSATLALGYSEGKSXSYSGDKYVDLDSVAPMKALVGVAWDD 684

RESULT 12

US-08-447-031A-2  
Sequence 2, Application US/08447031A  
Patent No. 5851794  
GENERAL INFORMATION:  
APPLICANT: GUSS, Bengt  
APPLICANT: HOOK, Magnus  
APPLICANT: JONSSON, Hans  
APPLICANT: LINDBERG, Martin  
APPLICANT: PATTI, Joseph  
APPLICANT: SIGNAS, Christer  
APPLICANT: SWITALSKI, Lech  
TITLE OF INVENTION: A COLLAGEN BINDING PROTEIN AS WELL AS  
TITLE OF INVENTION: ITS PREPARATION  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/447,031A  
FILING DATE: 22-MAY-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/861,804  
FILING DATE: 21-AUG-1992  
PRIOR APPLICATION DATA: WO PCT/SE91/00707  
FILING DATE: 22-OCT-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: SE 9003374-7

FILING DATE: 22-OCT-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: McGowan, Malcolm K.  
REGISTRATION NUMBER: 39,300  
REFERENCE/DOCKET NUMBER: 012889-006  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1183 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-447-031A-2  
Query Match 1.8%; Score 142.5; DB 1; Length 1183;  
Best Local Similarity 17.5%; Pred. No. 0.013;  
Matches 202; Conservative 180; Mismatches 419; Indels 355; Gaps 51;  
QY 505 DLWGAGLSDLVLI-----GPKSVRLYANTRDGFAGKGVVQSGD-ITLPPVPGADPRKL 556  
DB 32 DISSTNVTDLTVSPSKIEDGKTTVMPTDDKNG-----KIQNGDMIKVAMPTSGTVKI 85  
QY 557 VAFS-----DVLGS--GQAHLVEVSATKVTCPNLGRGRFGQPI-----TLPGFSQ----- 600  
DB 86 EGYSKTVPLTVKGEVQGVAVITPDGAT-IT-----FNDKVEKLSVSGFAFEVQG 135  
QY 601 -PATEFNPAQVYLAIDLDSGPTDLIYVHTNRL---DIFLKNKGNGFAEPVT-LRF----- 650  
DB 136 RNLQTNTSDDKVATITSGNKSTNVTYHKSEAGTSVFYKTDGMLPEDTTHVRWFLNIN 195  
QY 651 --PEGLRFPHDTQLOMADVQGLGVASLILSV--PMSPHHWRCDLTNMPML-----LN 700  
DB 196 NEKSYVSKDITIKQIQGGQQLDLSTLNINVTGTHSNYSYGSAITDFEKAPEGSKIITVD 255  
QY 701 EMNNMGVHHTLRYRSSQFWLDEKAAALTTGTQVPCYLPFPFHTLWQTEDETEDEISGNKL 760  
DB 256 NTQWTDVITPGYGSYNSFSINYKTKITNEQKEFVN---NSQAWYQEHGKEVNGKSF 312  
QY 761 VITLRYARGADGWRERFRGFGYVEQTDHQLAQNAPERTPPA-----LTKWYATGLP 815  
DB 313 NHTV-HNINANAGIEGTGKLVKQDK-----DTKAPTANYKFKLSKK---DGSV 360  
QY 816 VIDNALSTYWRDDQAFAGF-----SPRETTWQDNKDVPLTPEDDNSRYWF 861  
DB 361 VDNQKEIITDANGIANIKALPSGDYILKEIAPRPTFKDKKPYPTMKDITDQGYF 420  
QY 862 NRALKGQLLRS-----ELYGLDDSTNKHVPYVTVTEFRSQVRL 899  
DB 421 TTENAKAIEKTKDVSAQKWEQTKVKPTIYFKLYKQDDNQ-----ITPVDKAEIKKL 475  
QY 900 QHTDSRYPLVMSVSVSRNHYERIAADPQCSONITLSSDRFQPLKQLSVQVPRQOPA 959  
DB 476 E--DGTTKVTWSNLP-----NDKNGKAIKYLKVEYNAQGE-- 509  
QY 960 INLYPDTLPD-----KLLANS-----YDDQORQLRLTYQSSWHHLTN 997  
DB 510 -----DTTPEGYTKKENGVLVNTNTEKPIETTSISGEKVDKDDKDNQKRPKSVNLLAN 564  
QY 998 -NTRVRLGLPDSRSDIFTYGAENVPA--GGLNLELLSDKNSLIADDPREYL----- 1047  
DB 565 GEKVKTLP---DVTSETNWKYFKDLPKYDEGKKIEY-----TVTEHDVKDYTTDINGTT 615  
QY 1048 -----GQOKTAYT-----DQONTPTPTPTQALIAFTETTVFNQS-----TLS 1086  
DB 616 ITNKYTPGETSATVTKNWDDNNNDGKRPTKVELYQDGKATGKTAILNESNMTHTWT 675  
QY 1087 AFNGSIPSDKLSLTLEQ---AGYQQ-----TNLYFPRT---GEDKVVVAHHG 1127  
DB 676 GLDEKAKGQGVKVTBELTKVKGTYTHVDNMDGNLIVTNKYTPETTSISGEKVV----- 730  
QY 1128 YTDYGTAAQFWRPQKQSNLTQLTGKITLIWDANYCVVVQTRDAAGLTTSKDYWRFL----- 1183



Db 731 --DDKNDQGRPEKVS-----VLLADGE---KVKILDVTS---ETNWKYEFKDLPKYD 777  
Qy 1184 --TPVOLTIDINDNOHLITLIDALGRPITLRFMGTEGKMTGYSSPEKASFSPSDVNAABE 1241  
Db 778 EGKIEYTVTDHVDKDYTDINGTITNKY-----TPGETSAIV- 816  
Qy 1242 LKPLPVAQCQVYAPESMPVLVSKTFNRLBAQWQKLYNARIITEDGRICITLAYRRWQ 1301  
Db 817 -----TKNWDNNN-----QDKRTEIKVELYQ 840  
Qy 1302 SOKAIPOLISLLNGPRPLPHSLTLTDRYDHPDQRIQVQVPSDGFGRLLQAAARHEA 1361  
Db 841 DKGATCK-TALINES-----NNWHTWTGLD---EKAQQQVKYTVBELTKVKGYTHV- 890  
Qy 1362 GMAQRNEDGLSIIINVQHTENRWAVTGRTEYDNK-----GQPIRTYQ 1403  
Db 891 ---DNNDWGNLVNTKYPETTSISGEKVDKNDQDKRPEKVSNNLANGKVKTL 946  
Qy 1404 PYFLNDWRYVND SARQK-----EAYADTHVDPDIREIKVITAKGFRRTLTP 1454  
Db 947 VTSETNWKYBFDKLPKYDEKXKIEYTVTDEHVKDYTTDINGTITNKYTPGETSATVTN 1006  
Qy 1455 WFTVNEDENDTAAEVK 1470  
Db 1007 WDDNNQDKRPTK 1022

## RESULT 13

US-08-346-455B-38  
; Sequence 38, Application US/08346455B  
; Patent No. 573167  
; GENERAL INFORMATION:  
; APPLICANT: UNITED STATES OF AMERICA; DEPT.  
; APPLICANT: OF HEALTH AND HUMAN SERVICES  
; TITLE OF INVENTION: MOTILITY STIMULATING  
; TITLE OF INVENTION: PROTEIN USEFUL IN CANCER DIAGNOSIS AND  
; TITLE OF INVENTION: THERAPY  
; NUMBER OF SEQUENCES: 69  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MORGAN & PINNEGAN  
; STREET: 345 PARK AVENUE  
; CITY: NEW YORK  
; STATE: NEW YORK  
; COUNTRY: U.S.A.  
; ZIP: 10154  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy Disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: WordPerfect 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/346,455B  
; FILING DATE: 28-NOV-1994  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US95/06613  
; FILING DATE: 24-MAY-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/249,182  
; FILING DATE: 25-MAY-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 07/822,043  
; FILING DATE: 17-JAN-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: DOROTHY R. AUTH  
; REGISTRATION NUMBER: 36,434  
; REFERENCE/DOCKET NUMBER: 2026-4149PCT  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 758-4800  
; TELEFAX: (212) 751-6849  
; INFORMATION FOR SEQ ID NO: 38:  
; SEQUENCE CHARACTERISTICS:

LENGTH: 979  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: Unknown  
MOLECULE TYPE: protein  
HYPOTHETICAL: No  
ORIGINAL SOURCE:  
ORGANISM: Human  
STRAIN:  
INDIVIDUAL ISOLATE:  
DEVELOPMENTAL STAGE:  
HAPLOTYPE:  
TISSUE TYPE: Liver  
CELL TYPE:  
CELL LINE:  
ORGANELLE:  
FEATURE:  
NAME/KEY:  
LOCATION:  
IDENTIFICATION METHOD:  
OTHER INFORMATION: putative autotaxin  
OTHER INFORMATION: protein sequence from human liver  
US-08-346-455B-38  
Query Match 1.8%; Score 140; DB 1; Length 979;  
Best Local Similarity 18.9%; Pred. No. 0.015;  
Matches 139; Conservative 83; Mismatches 258; Indels 256; Gaps 34;  
Qy 418 WYSAORLGEIGSDAVTWKMQPLSVIPSLQSNASLVLDINGDGLDWDVI-----TGPGLRG 473  
Db 261 WITATKORGE-----SWN-----ILLVCHPSRAELITI-----LQWTLDPHERPSVA 305  
Qy 474 YHSQPDGSMTRFTPLNALPVEYTHPRAQLADLGA---GLSDLVLIIGPKSVRLYANTRD 530  
Db 306 FYSEQDPFSGHKHMPFGP---EMTNPLREMHKIVGQMDGLKQLKLRVCNVV-IFVETMD 361  
Qy 531 GFAK-----GKDVQSGDITLVPGA-----DPRKLVAFSDVLGSGQAH 569  
Db 362 GRCHMYRTEFLSNLYLTNNVDITL--VPGTLGRIRSKFSNNAKYDPKAIIA----- 409  
Qy 570 LVEVSATKVTCPNLPNGRGRFQPIITLPCFSOPATEFPNPAQVYLADLDGSGDPTDLYVHTN 629  
Db 410 -----NLTC-----KKPDQHPKP---YLKQ-----HLPRLHYANRR 438  
Qy 630 R-----LDIFLNKSGNGFAEVPVTLRFPPEGLRFDHTCOLQMDADVQGLG 671  
Db 439 RIEDIHLLVERRWVHARKPLDVYKKPSGNAFSETT-----AFDNKVNMQTVFVGYG 491  
Qy 672 -----VASLILSVPHMSPHWRCDLTNNKPMLLNENNNMGMVHTLYRVSQQFWLDEKAA 727  
Db 492 PTFKYTKVPPFENIELYNVMDLLGLKP-----APNNGTHGSLNHLRTNTFRPTMPE 545  
Qy 728 ALTTGQTP-VCYLPFP-----IHTLQOTETEDISGNKLVTT 763  
Db 546 EVTRPNYPGIMYLOSDFDLGCTCDDKVEPKNKLDELNKLRLTKGSTBERHLLYGDPAVL 605  
Qy 764 LR-----YARGAWDGRERBFRGFGVEQTDSDHOLAQNAPE-----RTPPALTK 807  
Db 606 YRTRYDILYHTDPESGYSEIFLMLPLWTSYTVSKAEVSSVPDHLTSCVRPDRVSPSPFSQ 665  
Qy 808 NWYATGLPVIDNALSTYWRDDQAFAG-----SP--RFTTWQONKQVLPTEDDNS 857  
Db 666 NCLA-----YKNDKQMSYGFLLFPYLLSSSPSEAKYDAFLVTNNVMPYPAFKRV 712  
Qy 858 RYWFENRAL-----KGQLARSELYGLDDSTNKHVPVTVTFEFSQVRRLQHTD 903  
Db 713 WNFQRLVVLKVIASERNGVNVISGIFDYDVGUHDTEDKIKQV-----EG 759  
Qy 904 SRYEVLWSSVSESNHYERIASDPQSONITLSSDRFGQPLKQLSVQYPRROOPAINLY 963  
Db 760 SSIPV-----PTHYSIITS-----CLDPTQADKCDGLSVSSFILPHR----- 799  
Qy 964 PDLPLDKLLANSYDDQQRQLRLTYQQSW---HHLTNNTRVRLGLPDDSTRSDIFTYGAENV 1021



```
FILE REFERENCE: 2026-4149US4
CURRENT APPLICATION NUMBER: US/09/483,831B
CURRENT FILING DATE: 2000-01-17
PRIOR APPLICATION NUMBER: 07/822,043
PRIOR FILING DATE: 1992-01-17
PRIOR APPLICATION NUMBER: 08/249,182
PRIOR FILING DATE: 1994-05-25
PRIOR APPLICATION NUMBER: 08/346,455
PRIOR FILING DATE: 1994-11-28
PRIOR APPLICATION NUMBER: 08/977,221
PRIOR FILING DATE: 1997-11-24
NUMBER OF SEQ ID NOS: 70
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 70
LENGTH: 979
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: Putative autotoxin protein sequence from human
OTHER INFORMATION: liver.
NAME/KEY: VARIANT
LOCATION: (860)
OTHER INFORMATION: Xaa at positions: 860, 889, 905, 911, 927, 937,
OTHER INFORMATION: 944, 950, 954, 967, and 975 represents an unknown
OTHER INFORMATION: or other amino acid.
US-09-483-831B-70

Query Match      1.8%; Score 140; DB 2; Length 979;
Best Local Similarity 18.9%; Pred. No. 0.015;
Matches 139; Conservative 83; Mismatches 258; Indels 256; Gaps 34;

Qy 418 WYRSNQRIGEIGSDAVTEKMQPLSVIPFSLQSNASLVLDINGDGLDWI-----TGPGLRG 473
Db 261 WITATKORGE-----SWN-----ILLVCCFSAEILTI-----LQMLTLPDHERPSVYA 305
Qy 474 YHSORPDGSTRFTPLNALPVEYTHPRAQLADLMGA---GLSDLVLIGPKSVRLYANTRD 530
Db 306 FYSEQPDPSGHKHPFGP---EMTNPLREMKIVGQLMDGLKQLKLRVCNV-IFVETMD 361
Qy 531 QPAK-----GKOVVQSGDITLPVPGA-----DPRKLVAFSVDVLGSGQAH 569
Db 362 GRCHMYRTEFLSNLTNNVDITL-VPGTLGRIRSKFSNNAKYDPKAIIA----- 409
Qy 570 LVEVSATKVTWPNLGRGRFGQPIITLPGFSQPATEFNPQAQVILADLGGSGPTDLYVHTN 629
Db 410 -----NLTC-----KKPDQHFKE---YLKQ---HLPKRLHYANNR 438
Qy 630 R-----LDIFLNKSGNGFAEPVTLFPPEGLREDHTCQLQMAADVQGLG 671
Db 439 RIEDIHLVVERHWVARPLDVIYKPKSGNAFSRETT-----AFDNKNSMQTVFVGYG 491
Qy 672 ----VASILISVPHMSPHHRCDLTNMKPMLNENNNMNVHHTLRYRSSQFWLDEKAA 727
Db 492 PTFKYKTKVPPENIELYNVMDLLGLKP-----APNNGTHGSLNHLRTNTPPTMPE 545
Qy 728 ALTTGQTP-VCYLPFP-----IHTLMQTEDEISGNKLVTT 763
Db 546 EVTRPNYFGIMYLSQDFDLGCTCDKVEPKNKLDELNKLRTKGTSTERRHLLYGDRAVL 605
Qy 764 LR-----YARGAWDGRERERFGYVEQTDHQLAQNAPE-----RTPPALTK 807
Db 606 YRTRYDILYHTDFESGYSEIFLMPILWTSYTSKQAEVSSVPDHLTSCVRPDVRVSPFSQ 665
Qy 808 NWYATGLPVIDNALSTEWDRDQAPAF-----SP--RFTTWQDNKDVPLTPEDDNS 857
Db 666 NCLA-----YKNDKQMSYGLFPFYLSSSPPEAKYDAFLVTNNVPMYPAPKRV 712
Qy 858 RYWFNRAL-----KGQLLRSELYGLDDSTNKHVPYTVTFEFSQVRLQHTD 903
Db 713 WNYFORVLVKVYASERNGVNWISGPIFDYDYGDLHDTEDKIKQYV-----EG 759
Qy 904 SRYPLWSSVVSRYHYERIASDPQCSQNTLSDRFGQPLKQLSVQYPRQOPAINLY 963
```

```
Db 760 SSIPV-----PTHYYSIITS-----CLDFTQPADKCDGFLSVSSFILPHR----- 799
Qy 964 PDTLPDKLLANSYDDQORQLRLTYQSSW--HHLTNNTVRVGLPDPSTRSDIFTYGAENV 1021
Db 800 PD---NBESCNSSED-----ESKWVEELMKMHTARVRDIEHLTSLDFFRKTSSSY 846
Qy 1022 PAGGLNLELLSDKNSL 1037
Db 847 P-----EILTKTYL 856
```

Search completed: December 16, 2005, 14:42:36  
Job time : 64 secs

*This Page Blank (uspto)*

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 16, 2005, 14:26:40 ; Search time 270 Seconds  
(without alignments)  
3851.662 Million cell updates/sec

Title: US-10-706-424-10

Perfect score: 7901

Sequence: 1 MQNSQDSITLSLPKGGGA.....WFTVNEDENTAAEVKKVKM 1474

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt\_05.80.\*

1: uniprot\_sprot.\*

2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	7901	100.0	1474	2	Q8GF99 PHOLU
2	7285.5	92.2	1475	2	Q7N7Y2 PHOLU
3	6065.5	76.8	1476	2	Q7N7Z0 PHOLU
4	6031.5	76.3	1476	2	Q93EP6 PHOLU
5	4627.5	58.6	1485	2	Q7N936 PHOLU
6	4595.5	58.2	1485	2	Q85153 PHOLU
7	3984.5	50.4	1433	2	Q8XP56 YERPR
8	3971	50.3	1428	2	Q9F9Z2 YENTR
9	3638	46.0	1489	2	Q6QDZ4 YERPS
10	3591	45.4	1481	2	Q693A4 YEREN
11	3569.5	45.2	1496	2	Q8ZAV4 YERPE
12	3569.5	45.2	1516	2	Q8D1P6 YERPE
13	3350.5	42.4	1401	2	Q3R8NE XENNE
14	2277.5	28.8	1447	2	Q4ZP56 PSESE
15	2274	28.8	1446	2	Q87X46 PSESM
16	1076	13.6	2439	2	Q4HWU2 GIBZE
17	837	10.6	591	1	VRP2 SALCH
18	835	10.6	591	1	P74846 SALTY
19	832	10.5	591	1	VRP2 SALTY
20	829	10.5	591	1	VRP2 SALEN
21	826.5	10.5	593	1	VRP2 SALDU
22	826.5	10.5	593	2	Q4L1X7 9ENTR
23	709.5	9.0	563	2	P72421 9ENTR
24	676	8.6	283	2	Q8KSN9 XENNE
25	651.5	8.2	2031	2	Q3XEC6 BURPS
26	502	6.4	1806	2	Q62N48 BURMA
27	457.5	5.8	166	2	Q6WCC1 9ENTR
28	386	4.9	2217	2	Q8TP72 METAC
29	330.5	4.2	2379	2	Q72U39 LEPIC
30	326.5	4.1	221	2	Q58FC3 SALGL
31	292	3.7	2306	2	Q8F107 LEPIN

32 289.5 3.7 2554 2 Q72QR5 LEPIC Q72qr5 leptospira  
33 279.5 3.5 2321 2 Q8F5B9 LEPIN Q8f5b9 leptospira  
34 241.5 3.1 1826 2 Q987Z7 RHIL0 Q987z7 rhizobium 1  
35 202 2.6 2370 2 Q82RE3 STRAW Q82re3 streptomyce  
36 199 2.5 2060 2 Q6Q8B0 STALU Q6q8b0 staphylococ  
37 181.5 2.3 2364 2 Q82R58 STEAM Q82r58 streptomyce  
38 180 2.3 1976 2 Q8D4R5 VIBVU Q8d4r5 vibrio vuln  
39 177 2.2 795 2 Q7NFS7 GLOVI Q7nfs7 gloeobacter  
40 174.5 2.2 1881 2 Q8TJS7 METAC Q8tjs7 methanosarc  
41 172.5 2.2 1404 2 Q7AH10 ECO57 Q7ah10 escherichia  
42 172.5 2.2 1404 2 Q8XED9 ECO57 Q8xed9 escherichia  
43 171.5 2.2 2566 2 Q8TS87 METAC Q8tse7 methanosarc  
44 170 2.2 593 2 Q8GP37 PSEAE Q8gp37 pseudomonas  
45 170 2.2 2082 2 Q9ACP4 STRCO Q9acp4 streptomyce

## ALIGNMENTS

### RESULT 1

Q8GF99 PHOLU PRELIMINARY; PRT; 1474 AA.  
AC Q8GF99;  
DT 01-MAR-2003 (Tremblrel. 23, Created)  
DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)  
DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)  
DE TcdB2.  
GN Name=tcdB2;  
OS Photorhabdus luminescens (Xenorhabdus luminescens).  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Photorhabdus.  
OX NCBI\_TaxID=29488;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=W14;  
RX MEDLINE=21185117; PubMed=11286884; DOI=10.1016/S0966-842X(01)01978-3;  
RA Waterfield N.R., Bowen D.J., Fetherston J.D., Perry R.D.,  
RA ffrench-Constant R.H.;  
RT "The tc genes of Photorhabdus: a growing family.";  
RL Trends Microbiol. 9:185-191(2001).  
RN [2]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=W14;  
RX MEDLINE=22454990; PubMed=12564983; DOI=10.1016/S0966-842X(02)02463-0;  
RA Waterfield N.R., Daborn P.J., ffrench-Constant R.H.;  
RT "Genomic islands in Photorhabdus.";  
RL Trends Microbiol. 10:541-545(2002).  
DR EMBL: AF346500; AA017202.1; -; Genomic\_DNA.  
DR GO: GO:0005737; C:cytoplasm; IEA.  
DR GO: GO:0008305; C:integrin complex; IEA.  
DR GO: GO:0007160; P:cell-matrix adhesion; IEA.  
DR InterPro: IPR000413; Integrin alpha.  
DR InterPro: IPR003284; Sal\_SpVb.  
DR Pfam: PF01839; EG-GAP; 2.  
DR Pfam: PF03534; SpvB; 1.  
DR PRINTS: PR01341; SALSPVBPROT.  
SQ SEQUENCE 1474 AA; 166338 MW; 9118D4E4914683DD CRC64;

Query Match 100.0%; Score 7901; DB 2; Length 1474;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1474; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MQNSQDSITLSLPKGGGAITGMGEALTPTGPDGMAALSPLPISAGRGPVAPFTLNN 60  
Db 1 MQNSQDSITLSLPKGGGAITGMGEALTPTGPDGMAALSPLPISAGRGPVAPFTLNN 60  
Qy 61 SGAGNSPFLGWCDCNVMTRRRTHFGVPHYDEDTFTFLGPEGEVLVADQPRDESTLQGIN 120  
Db 61 SGAGNSPFLGWCDCNVMTRRRTHFGVPHYDEDTFTFLGPEGEVLVADQPRDESTLQGIN 120  
Qy 121 LGATFTVTGYSRLESFHSRLEYWQPKTKGTDFWLYSPDQVHLLGKSPQARSNPQ 180  
Db 121 LGATFTVTGYSRLESFHSRLEYWQPKTKGTDFWLYSPDQVHLLGKSPQARSNPQ 180

QY	181	TTQTAAWLLLEASVSSRGEQIYYQVRAEDDTCBEADEITHHLQATAQRYLHIVYGNRTAS	240
DB	181	TTQTAAWLLLEASVSSRGEQIYYQVRAEDDTCBEADEITHHLQATAQRYLHIVYGNRTAS	240
QY	241	ETLPGLDGSAPSQADLFLYLVDYFGERSNNLKTTPAFSTTGSWLCQDRFSRYEYGFPIR	300
DB	241	ETLPGLDGSAPSQADLFLYLVDYFGERSNNLKTTPAFSTTGSWLCQDRFSRYEYGFPIR	300
QY	301	TRRLCQVLMYHHLQALDLSKITEHNGPTLVSRILNLYDESAIASTLVFVRVCHQDGNV	360
DB	301	TRRLCQVLMYHHLQALDLSKITEHNGPTLVSRILNLYDESAIASTLVFVRVCHQDGNV	360
QY	361	VTLPPLLELAYQDFSPRHHAHQMDVLANFNALQRMQLVDLKGEGPLGLLYQDKGAWYR	420
DB	361	VTLPPLLELAYQDFSPRHHAHQMDVLANFNALQRMQLVDLKGEGPLGLLYQDKGAWYR	420
QY	421	SAQRLGEIGSDAVTWKMQPLSVIPSLQSNASLVNDINGDQQLDWITGPGLRGYHSQRPD	480
DB	421	SAQRLGEIGSDAVTWKMQPLSVIPSLQSNASLVNDINGDQQLDWITGPGLRGYHSQRPD	480
QY	481	GSWTRFTPLNALPVEYTHPRAQADLADLMGAGLSDLVLIGPKSVRLYANTRDGPAPKGDVVQ	540
DB	481	GSWTRFTPLNALPVEYTHPRAQADLADLMGAGLSDLVLIGPKSVRLYANTRDGPAPKGDVVQ	540
QY	541	SGDITLTPVPGADPRKLVAFSDVLGSGOAHLEVSATKVTCPNIGRGRFGQPIITLPGFSQ	600
DB	541	SGDITLTPVPGADPRKLVAFSDVLGSGOAHLEVSATKVTCPNIGRGRFGQPIITLPGFSQ	600
QY	601	PATEFNPAAVYLDLQSGPTDILYVHTNRDLFLANKSGNGFAPVTLRFPFGLRFDHTC	660
DB	601	PATEFNPAAVYLDLQSGPTDILYVHTNRDLFLANKSGNGFAPVTLRFPFGLRFDHTC	660
QY	661	QLOWADVQGLGVASLILSVPHMSPHWRCDLTNMKPWLNMNNMNGVHTLRYRSSQP	720
DB	661	QLOWADVQGLGVASLILSVPHMSPHWRCDLTNMKPWLNMNNMNGVHTLRYRSSQP	720
QY	721	WLDEKAAALTGTGTPVCYLPPTHTLMQTEDEISGNKLVITTLRYARGADGRCRERFRG	780
DB	721	WLDEKAAALTGTGTPVCYLPPTHTLMQTEDEISGNKLVITTLRYARGADGRCRERFRG	780
QY	781	FGYVEQTDSHQLAQGNAPERTPALTKWYATGLPVIDNALSTEYMRDQAFAGFSRPT	840
DB	781	FGYVEQTDSHQLAQGNAPERTPALTKWYATGLPVIDNALSTEYMRDQAFAGFSRPT	840
QY	841	TWQDNKDVPLTPEDDNRSYFNALKGQLLRSELYGLDDSTNKHVPYTVTFEFSQVRRLQ	900
DB	841	TWQDNKDVPLTPEDDNRSYFNALKGQLLRSELYGLDDSTNKHVPYTVTFEFSQVRRLQ	900
QY	901	HTDSRYPVLASVSVESRNYHYERTASDPQCSQNTLSSDRFGQPLKOLSVQYPRQOPAI	960
DB	901	HTDSRYPVLASVSVESRNYHYERTASDPQCSQNTLSSDRFGQPLKOLSVQYPRQOPAI	960
QY	961	NLYPDTLPDKLLANSYDDQQRQLRLTYQQSSWHHLTNNTVRLGLPDLSTRSDIFTYGAEN	1020
DB	961	NLYPDTLPDKLLANSYDDQQRQLRLTYQQSSWHHLTNNTVRLGLPDLSTRSDIFTYGAEN	1020
QY	1021	VPAGGLNLELLSDKNSLIADKPREYLGQOKTAYTDGQNTTPTQTPROALIAFTETTVF	1080
DB	1021	VPAGGLNLELLSDKNSLIADKPREYLGQOKTAYTDGQNTTPTQTPROALIAFTETTVF	1080
QY	1081	NOSTLSAFNGSIPSDKLSLTLEQAGYQOTNYLPRPTGEDKVVVAHGVTYDGTAAQFWRP	1140
DB	1081	NOSTLSAFNGSIPSDKLSLTLEQAGYQOTNYLPRPTGEDKVVVAHGVTYDGTAAQFWRP	1140
QY	1141	QKQNTQLTGKILIMDANYCVVVQTRDAAGLTTSKAYDWRFLTPVQLTDINDNQHILITL	1200
DB	1141	QKQNTQLTGKILIMDANYCVVVQTRDAAGLTTSKAYDWRFLTPVQLTDINDNQHILITL	1200
QY	1201	DALGRPTTLRPWGTENGKMTGYSSPEKASPPSDVNAAIELKKPLPVAQCQVYAPESWM	1260
DB	1201	DALGRPTTLRPWGTENGKMTGYSSPEKASPPSDVNAAIELKKPLPVAQCQVYAPESWM	1260
QY	1261	PVLSQKTFNRLAEQDWQKLYNARIITEDGRICITLAYRRVWQSKAIPQLISLNNNGPRLP	1320
DB	1261	PVLSQKTFNRLAEQDWQKLYNARIITEDGRICITLAYRRVWQSKAIPQLISLNNNGPRLP	1320
QY	1321	PHSLTLTTORYDHDPEQRIROQVVFSDGFGRLLOAAARHEAGMARQNRNEDGSLIINVQHT	1380
DB	1321	PHSLTLTTORYDHDPEQRIROQVVFSDGFGRLLOAAARHEAGMARQNRNEDGSLIINVQHT	1380
QY	1381	ENRMWVGTGVNDKQPIRTYQPYFLNDWRVYVNSDSARQEKAYADTHVYDPIGRIKV	1440
DB	1381	ENRMWVGTGVNDKQPIRTYQPYFLNDWRVYVNSDSARQEKAYADTHVYDPIGRIKV	1440
QY	1441	ITAKGFRFTLTPFWFTVNEDENDTAAEVKKVKM	1474
DB	1441	ITAKGFRFTLTPFWFTVNEDENDTAAEVKKVKM	1474
RESULT 2			
Q7NTY2_PHOLL			
ID	Q7NTY2_PHOLL	PRELIMINARY;	PRT; 1475 AA.
AC	Q7NTY2_	PHOLL	
DT	01-MAR-2004	(TREMBLrel. 26, Created)	
DT	01-MAR-2004	(TREMBLrel. 26, Last sequence update)	
DT	01-MAR-2004	(TREMBLrel. 26, Last annotation update)	
DE	Insecticidal toxin complex protein TcdB2.		
GN	Name=cdb2; OrderedLocName=plu0969;		
OS	Phototaxobacterium luminescens (subsp. laumondii).		
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;		
OC	Enterobacteriaceae; Phototaxobacterium.		
OX	NCBI_TaxID=141679;		
RN	[1]		
RP	NUCLEOTIDE SEQUENCE.		
RC	STRAIN-T01;		
RX	MEDLINE=22957637; PubMed=14528314; DOI=10.1038/nbt886;		
RA	Duchaud E., Rubnik C., Frangeul L., Buchrieser C., Givaudan A.,		
RA	Taurat S., Bosc S., Bouraux-Bude C., Chaudier M., Charles J.-P.,		
RA	Dassa E., Derzelle S., Derzelle S., Freysinet G., Gaudriault S.,		
RA	Medigue C., Lanois A., Powell K., Sigulier P., Vincent R., Wingate V.,		
RA	Zouine M., Glaser P., Boemare N., Danchin A., Kunst F.,		
RT	"The genome sequence of the entomopathogenic bacterium Phototaxobacterium luminescens.";		
RL	Nat. Biotechnol. 21:1307-1313(2003).		
DR	EMBL; BX571862; CAE13264.1; -; Genomic_DNA.		
DR	Photoliat; plu0969; -		
DR	GO; GO:0005737; C:cytoplasm; IEA.		
DR	GO; GO:0008305; C:integrin complex; IEA.		
DR	GO; GO:0007160; P:cell-matrix adhesion; IEA.		
DR	InterPro; IPR000413; Integrin_alpha.		
DR	InterPro; IPR003284; Sal_SpVb.		
DR	Pfam; PF01839; FG-GAP; 2.		
DR	Pfam; PF03534; SpvB; 1.		
DR	PRINTS; PR01341; SALSVPBPROT.		
DR	Complete proteome.		
QW	SEQUENCE	1475 AA; 165779 MW; 24945174 CDADAD22 CRC64;	
Query Match 92.2%; Score 7285.5; DB 2; Length 1475;			
Best Local Similarity 91.7%; Pred. No. 0;			
Matches 1352; Conservative 55; Mismatches 67; Indels 1; Gaps 1;			
QY	1	MONSQDSITELSLPKGGGAIQWGEALTPGPDGMAALSPLPI SAGRGYAPFTLNYN	60
DB	1	MONSQDSITELSLPKGGGAIQWGEALTPGPDGMAALSPLPI SAGRGYAPFTLNYN	60
QY	61	SGAGNPFGLGWCNVTIRRRTHFGVPHYDETDFTLGPGEVNVADQDRDESTLOGIN	120
DB	61	SGAGNPFGLGWCNVTIRRRTHFGVPHYDETDFTLGPGEVNVADQDRDESTLOGIN	120
QY	121	LGATFTVTGYSRLESHFSLRYWQPKTKGTDFWLIYSPDGVHLLGKSPQARISNPSQ	180
DB	121	LGATFTVTGYSRLESHFSLRYWQPKTKGTDFWLIYSPDGVHLLGKSPQARISNPSQ	180
QY	181	TTQTAAWLLLEASVSSRGEQIYYQVRAEDDTCBEADEITHHLQATAQRYLHIVYGNRTAS	240



Db 181 TTQTQAWLLEASVPHGSGQIYYQYRAEDNLDCEITNEITLHPQATAQRVLHIVYGNRTAS 240  
Qy 241 ETLGCLDGSAPSQADWLFPYLPVDYGERSNLKTTPAFSTTGSWLCRODRPGRYEGFEIR 300  
Db 241 ETLGCLNAGRAPSQADWLFPYLPVDYGERSNLKMPPAFATGNWLCRODRPGRYEGFALR 300  
Qy 301 TRRLCROVLMVHLLQALDSKITEHNGPTLVSRLLIINYDESAIATSLVPRVRGHEQDGNV 360  
Db 301 TRRLCROVLMVHLLQALDNKIKENHNGPTLVSRLLIINYDESAIATSLVPRVRGHEQDGT 360  
Qy 361 VTLPLELAYQDFSPRHHAHQMDVLNFAIQRWQLVDLKGEGPLGLLYQDKGAMWYR 420  
Db 361 VTLPLELAYQDFSPQHNTDQPMVDLANFNAIQRWQLVDLKGEGPLGLLYQDKNAMWYR 420  
Qy 421 SAQRIGEGSDAVTWKQKQPLSVIPSLQSNASLVDINGDGLDWITTPGLRGHYSORPD 480  
Db 421 SAQRIGEGSDAVTWKQKQPLSVIPSLQSNASLVDINGDGLDWITTPGLRGHYSORPD 480  
Qy 481 GSWTRFTPLNALPVEYTHPRQAQLADLMGAGLSDLVLIGPKSVRLYANTRDGFAGKQVQ 540  
Db 481 GSWTRFTPLNALPVEYTHPRQAQLADLMGAGLSDLVLIGPNSVRLYANTRDGFAGKQVQ 540  
Qy 541 SGGITLPPVPGADPRKLVAFSDVLSGGQAHLEVSATKVTCPNLRGRGFGQPITLPGFSQ 600  
Db 541 SGGITLPPVPGADPRKLVAFSDVLSGGQAHLEVSATKVTCPNLRGRGFGQPITLPGFSQ 600  
Qy 601 PATEFPAQVYLAIDLGGPTDLIYVHTNRDLDFLNKSGNGFABPVTLPPEGLRFDHTC 660  
Db 601 PEALFPAQVYLAIDLGGPTDLIYVHTNRDLDFLNKSGNGFABPVTLPPEGLRFDHTC 660  
Qy 661 QLOWADVOGLVASILSVPHMSPHWRCDLTNNKPMLENNMNGVHHTLYRSSQF 720  
Db 661 QLOWADVOGLVASILSVPHMTPHWRCDLTSAKPMLENNMNGVHHTLYRSSQF 720  
Qy 721 WLDEKAAALTTGQTPVCVLPPIHTLMQTEDEISGNKLVTLRYARGAGWDGRERPRG 780  
Db 721 WLDEKAAALVAGQTPVCVLPPIHTLMQTEDEISGNKLVTLRYARGAGWDGRERPRG 780  
Qy 781 FGYYEQTDSHQAQGNAPERTPPALTKRWYATGLPVIDNALSTEWYR-DDQAQAGFSPRF 839  
Db 781 FGYYEQTDSHQAQGNASERTPPAMTKWYATGLPVIDNTLSTEWYRGNQAFAGFSPRF 840  
Qy 840 TTQCDNDKVPITPDDNSRYWPNFALKQGLRSELYGLDSTNKHVPYTVTFPSQVRL 899  
Db 841 TTQCDNDKVPITPDDNSRYWPNFALKQGLRSELYGLDSTNKHVPYTVTFPSQVRL 900  
Qy 900 QHTDSRYPLWSSVYVESRYHRIASDPQCSQITLSSDRFGQPLKQSVQYPRRQOPA 959  
Db 901 QOANQYPLWSSVYVESRYHRIASDPQCSQITLSSDRFGQPLKQSVQYPRRQOPA 960  
Qy 960 INLYPDTLPDKLANSYDDOQRLRLTYQSSWHHTLNTVTVLGLPDSRSDIFTYGA 1019  
Db 961 ISPYEDTLPEDELLANSYDDOQRLRLTYQSSWHHTLNTVTVLGLPDSRSDIFTYGA 1020  
Qy 1020 NVPAGLNLELLSNKSLIADKREYLGQKATYTDQNTTPTQTPRQALIAFTETTV 1079  
Db 1021 NVPAGLNLELLSNKSLIADKREYLGQKATYTDQNTTPTQTPRQALIAFTETTV 1080  
Qy 1080 FNQSTLSAFNGSIPSDKLSLTLEAGYQOYNLPRTGDKVWAHGYTDYGTAAQFWR 1139  
Db 1081 FSQSTLSAFNGSIPSAQSLTLEAGYQOYNLPRTGDKVWAHGYTDYGTAAQFWR 1140  
Qy 1140 POKQNTQLTGKITLINDANYCVVQTRDAAGLTTSYAKYDWRFTFPVQLTDINDNOHLIT 1199  
Db 1141 POKQNTQLTGKITLINDANYCVVQTRDAAGLTTSYAKYDWRFTFPVQLTDINDNOHLIT 1200  
Qy 1200 LDALGRPTTLFPWGTENGKMTGYSPKASPSPSDVNAALTKPLPVACQVYAPESW 1259  
Db 1201 LDALGRPTTLFPWGTENGKMTGYSPKASPSPSDVNAALTKPLPVACQVYAPESW 1260  
Qy 1260 MPVLSQKTFNRLAQDQWQKLNARILITDGRICTLAYRRWQSOAKAPOLLSLNNGPRL 1319  
Db 1261 MPVLSQKALNRLAQDQWQKLNARILITDGRICTLAYRRWQSOAKAPOLLSLNNGPRL 1320

Qy 1320 PPHSLTUTTRYDHDHPQOIRQQVVFSDGFRLLQAAARHAGMARQNRNEDGSLIINVQH 1379  
Db 1321 PPHSLTUTTRYDHDHPQOIRQQVVFSDGFRLLQAAARHAGMARQNRNEDGSLIINVQH 1380  
Qy 1380 TENRWAVTGRTEYDNKQPIRTYQYPLNDWRVYNSDSARQEKAYADTHVYDPIGRK 1439  
Db 1381 TENRWAVTGRTEYDNKQPIRTYQYPLNDWRVYNSDSARQEKAYADTHVYDPIGRK 1440  
Qy 1440 VITAKGFRRTLTFTFWTNEDENDTAAEVKKVKH 1474  
Db 1441 VITAKGFRRTLTFTFWTNEDENDTAAEVKKVKH 1475  
RESULT 3  
Q7N7Z0 PHOLL PRELIMINARY; PRT; 1476 AA.  
AC Q7N7Z0;  
DT 01-MAR-2004 (TREMREL. 26, Created)  
DT 01-MAR-2004 (TREMREL. 26, Last sequence update)  
DT 01-MAR-2004 (TREMREL. 26, Last annotation update)  
DE Insecticidal toxin complex protein TcdB1.  
GN Name=tcdB1; OrderedLocusNames=plu0961;  
OS Photorhabdus luminescens (subsp. laumondii).  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Photorhabdus.  
OX NCBI\_TaxID=141679;  
RN (1)  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=TT01;  
RX MEDLINE=22957627; PubMed=14528314; DOI=10.1038/nbt886;  
RA Duchaud E., Rusniok C., Frangeul L., Buchrieser C., Givaudan A.,  
RA Taurat S., Bocs S., Bouraux-Eude C., Chandler M., Charles J.-F.,  
RA Dassa E., Derosé R., Berzelle S., Freysinet G., Gaudriault S.,  
RA Meunier C., Lanois A., Powell K., Siguer P., Vincent R., Wingate V.,  
RA Zouine M., Glaser P., Boenare N., Danchin A., Kunat F.,  
RT "The genome sequence of the entomopathogenic bacterium Photorhabdus  
RT luminescens.",  
RL Nat. Biotechnol. 21:1307-1313(2003).  
DR EMBL; BX571862; CAB13256.1; -, Genomic\_DNA.  
DR PhotoList; plu0961; -.  
DR GO; GO:0005737; C:cytoplasm; IEA.  
DR GO; GO:0008305; C:integral complex; IEA.  
DR GO; GO:0007160; P:cell-matrix adhesion; IEA.  
DR InterPro; IPR000413; Integrin alpha.  
DR InterPro; IPR003284; Sal\_SpV8.  
DR Pfam; PF01839; FG-GAP; 2.  
DR Pfam; PF03534; SpvB; 1.  
DR PRINTS; PR01341; SALSPVBPROT.  
KW Complete proteome.  
SQ SEQUENCE 1476 AA; 165737 MW; 06AFE0EB945B5728 CRC64;  
Query Match 76.8%; Score 6065.5; DB 2; Length 1476;  
Best Local Similarity 76.3%; Pred. No. 0;  
Matches 1126; Conservative 120; Mismatches 221; Indels 9; Gaps 3;  
Qy 1 MNSQDSFISITELSLPKGGGALTGMGEALTPTGPGMAALSPLPISAGRYAPFTLNYN 60  
Db 1 MNSQTSFVAELSLPKGGALTGMGEALTPTGPGMAALSPLPISAGRYSPSITLNS 60  
Qy 61 SGAGNSPFLGWDNCNVMTIRRRTHFGVPHYDEDTFLGPGGEVLVA-----DQPRDES 114  
Db 61 SGAGNSPFLGWDNCNVMTIRRRTHFGVPHYDEDTFLGPGGEVLVAALNENGQADIRSES 120  
Qy 115 TLQGINLGAFTTVTGYSRLSHESFSLRYMQTKTGKDFWLIYSPGQVHLGKSPQAR 174  
Db 121 SLQGINLGEIFTTVGYSRLSHESFSLRYMQTKTGKDFWLIYSPGQVHLGKSPQAR 180  
Qy 175 ISNPSQTTQAWLLEASVSRGEGQIYYQYRAEDDTCCEADEITHHQAQRYLHIVY 234  
Db 181 ISNPLNVSQTAQWLLLEASVSHGEGQIYYQYRAEDTNCETDEFTAHFNATVQYLAQVHY 240  
Qy 235 GNRTASTETPLGCLDGSAPSQADWLFPYLPVDYGERSNLKTTPAFSTTGSWLCRODRPFSRYE 294

```
Db 241 GNLTADEVPTLNGDDPLKSGMLFCLVDFYGERKNSLSIPEPPKASSLWLCRDQDFSRYE 300
Qy 295 YGEIETRLRCQVLAHYHLQALDSKITEHNGPTLVSRILNLYDESAIASTLVFVRVGH 354
Db 301 YGPELTRRLCRLQILFHRILQTLGSGQAKGDEPALVSRLLTLDYDANAVISTLVSRRIH 360
Qy 355 EQGNNVVTLPPLSLAYQDFSPRHAHWQPMQDVLNANFNAIORMQLVDLKGELPGLLYQDK 414
Db 361 EDNNTVISLPPLESLAYQDFPEPEQKARQSMQDVLNANFNAIORMQLVDLKGEGVGVLYQDR 420
Qy 415 GAWWYSAQRGLBGISDAVWEMKQPLSVIPISQSNASLVNDINGDQGLDWVITPGRLGY 474
Db 421 NGWWYSAQRQAQGEENNAVTGKWLPLPITPALQDNASLMDINGDQGLDWVITPGRLGY 480
Qy 475 HSQRPPGSWTRFPFANALPVEYTHPRAQLADLNGAGLSDLVLGPKSVRLYANTRGPAK 534
Db 481 HSQHPDGSWTRFPFALPIEYSHPRQAQLADLNGAGLSDLVLGPKSVRLYANNRDFTQ 540
Qy 535 GKDVQSGDITLPPVPGADPRKLVAFSDVLGSGQAHLEVEVSATKVTQWPNLGRGRFGQPI 594
Db 541 GRDVQSGDITLPPGADARKLVAFSDVLGSGQAHLEVEVSATQVTCWPNLCHGRFGQPI 600
Qy 595 LPGFSQAPATEFNAQVYLADLQSGPTDLIYVHTNRLDIFLNKSGNGFAPBPVTLRPEGL 654
Db 601 LPGFSQSDNFNDRVHLADLQSGPADLIYVHTDRLEIFKNESGNSFAKPFVTLRPPDGL 660
Qy 655 RPDHTCOLQWADVQGLGVASLIISVPHMSPHWRCDLTNKPWLLNEMNNMNGVHTLRY 714
Db 661 RPDHTCOLQWADVQGLGVASLIISVPHMAPHWRCDLTNAKPWLLSEMNNMNGAHTLHY 720
Qy 715 RSSSQFWLDBKAAALTGTGQPCVCLPPIHTLMQTEDEISGNKLVTLTRYARGAWDGR 774
Db 721 RSSSQFWLDBKAAALTGTGQPCVCLPPIHTLMQTEDEISGNKLVTLTRYARGAWDGR 780
Qy 775 EREFRGFGYVEQTDSHOLAQGNAPERTPPALTKNWYATGLFVIDNALSTEYMR-DDQAFA 833
Db 781 EREFRGFGYVEQTDSHOLAQGNAPERTPPALTKNWYATGYPEVDNTLSAGYMRGDKQAF 840
Qy 834 GFSRFTTQDNKDVPITPBDNSRWYFNALGKQLLSRLYGLDSTNKHVYTVYTFPR 893
Db 841 GFTFRFTWKEGKDVATPENNDDNLYWFNALGKQLLSRLYGLDSEQQNIPTVYTFESR 900
Qy 894 SOVRLQHTDSRYPLVWSSVVERSNHYHRIASDPQCSQNTILSSDRFGQPLKOLSQVYP 953
Db 901 PQVRLQDGTASPLVWSSVVERSNHYHRIIGDPCQCNQDITLSSDQFGQPLKQVSNQYP 960
Qy 954 RROQPAINLYPDLPLDKLANSYDDQORQLRLTYQSSMHHLTNNVTVRVLGLPDRSDI 1013
Db 961 RRNQPTTNPYPDLPTDLTFASSYDDQQLRLTLCQSSMHHLTGNELRVLGLPDGTRSDA 1020
Qy 1014 FTYGAENVAGGNLELLSDKNSLIADDKPREYLGQOKTAYTQGNTPLOTTQROLIA 1073
Db 1021 FTYDAKQVPVDGLNLEALCAENSLIADDKPREYLNQORTFTYDGKNOAPLEIPTQROLIA 1080
Qy 1074 FTETTVNOSTLAFNGSIPSDKLSLTLEQAGYQOONTNLPRTGEDKVVAAHGYTDYGT 1133
Db 1081 FTETAVLTESLLSAFDGGITPDELPGILTQAGYQOEPYLPRTGENKVVWARGYTDYGT 1140
Qy 1134 AAQFWRPQKQNTQLTGKTLIWDANYCVVVQVTRDAAGLTTSAKYDWRFLTPVQLTDIND 1193
Db 1141 EAQFWRPVQARNVTLTGKTLIWDANYCVVQVTRDAAGLTTSAKYDWRFLTPVQLTDIND 1200
Qy 1194 NQHLITLDALGRFPIITLFWGTENGKMTGYSSPKASFPSPDVNAAITELKKPLFVACQV 1253
Db 1201 NVHLITLDALGRFPIITLFWGTENGKMTGYSSPKASFPSPDINTAIALTGPLFVACQV 1260
Qy 1254 YAPESNMPVLISQTFNPLASQDWKLYNARIITEDGRICTIAYRRWQSKAPOLISLL 1313
Db 1261 YAPDSNMPLFSQTFNPLASQDWKLYNARIITEDGRICTIAYRRWQSKAPOLISLL 1320
Qy 1314 NNGPRLPHSLITLTDYRDHDPQOIRQVQVFSDFGRLGLQAARHAGMARORNEGSL 1373
```

```
Db 1321 TNSIGLPPHNLTLTPDRYDRDSQQIHQVAFSDGFRLLQASVRHEAGAMQRNODGSL 1380
Qy 1374 IINVOHTENRWAVTGRTEYDNKGQPIRTYQPYFLNDWRVYVNSDSARQEKAEVADTHVYDP 1433
Db 1381 VYTMEDTKRWAVTGRTEYDNKGQPIRTYQPYFLNDWRVYVNSDSAR--KGAVADTHIYDP 1438
Qy 1434 IGREIKVITAKGFRRTLTFTPMFTVNEDENDTAAEV 1469
Db 1439 IGREIRVITAKGWLRSQSYFPWFVTVSEDENDTAAV 1474

RESULT 4
Q93EP6 PHOLU
ID Q93EP6 PHOLU PRELIMINARY; PRT; 1476 AA.
AC Q93EP6
DT 01-DEC-2001 (T-EMBLrel. 19, Created)
DT 01-DEC-2001 (T-EMBLrel. 19, Last sequence update)
DT 01-MAR-2004 (T-EMBLrel. 26, Last annotation update)
DE TcdB1.
GN Name=tcdB1;
OS Photorhabdus luminescens (Xenorhabdus luminescens).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Photorhabdus.
OX NCBI_TaxID=29488;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=W14;
RX MEDLINE=21185117; PubMed=11286884; DOI=10.1016/S0966-842X(01)01978-3;
RA Waterfield N.R., Bowen D.J., Fetherston J.D., Perry R.D.,
RA french-Constant R.H.;
RT "The tc genes of Photorhabdus: a growing family.";
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=W14;
RX MEDLINE=22454990; PubMed=12564983; DOI=10.1016/S0966-842X(02)02463-0;
RA Waterfield N.R., Daborn P.J., french-Constant R.H.;
RT "Genomic islands in Photorhabdus.";
RX Trends Microbiol. 10:541-545(2002).
DR EMBL; AF346500; AAL18487.1; -; Genomic_DNA.
DR GO; GO:0005737; Cytoplasm; IEA.
DR GO; GO:0008305; C:integrin complex; IEA.
DR GO; GO:0007160; P:cell-matrix adhesion; IEA.
DR InterPro; IPR000413; Integrin alpha.
DR InterPro; IPR003284; Sal_SpYB.
DR Pfam; PF01839; FG-GAP; 2.
DR Pfam; PF03534; SpvB; 1.
DR PRINTS; PR01341; SALSPVBPROT.
SQ SEQUENCE 1476 AA; 165138 MW; 8B6AC3D109911995 CRC64;

Query Match 76.3%; Score 6031.5; DB 2; Length 1476;
Best Local Similarity 75.9%; Pred. No. 0;
Matches 1120; Conservative 122; Mismatches 224; Indels 9; Gaps 3;

Qy 1 MNSQSFSTELSLPKGGGAIATGMEALTPGPDGMAALSLPLISAGRGYAPAFILNYN 60
Db 1 MNSQSFSTELSLPKGGGAIATGMEALTPGPDGMAALSLPLISAGRGYAPAFILNYN 60
Qy 61 SGAGNSPFGLGWDCNMTIRRTTHFGVPHYDETDFTLGPGEVILVVA-----DQPRDES 114
Db 61 SGTGNSPFGLGWDCGWAIRRTTSTGPNYDETDFTLGPGEVILVVALEAGQADIRSES 120
Qy 115 TLQGINLGATFTVTGVRSLRLESHFSRLYWPQTKTKTDFTWLIYSDGQVHLLGKSPQAR 174
Db 121 SLQGINLGATFTVTGVRSLRLESHFSRLYWPQTKTKTDFTWLIYSDGQVHLLGKSPQAR 180
Qy 175 ISNPSQTTQTAOWLLASVSSRGEQIYYOVRADDDTGCEADETHLQATAQRYLHVY 234
Db 181 ISNPLNVNQAQWLLASVSSRGEQIYYQVRADEAGCETDELAHPSATVQRYLQTVHY 240
Qy 235 GNRATSETLPLGDSAPSQADWLFLYVDFYGERSNLKTTPAFSTTGSWLCRODRFSRYE 294
Db 241 GNLTADEVPTLNGDDPLKSGMLFCLVDFYGERKNSLSIPEPPKASSLWLCRDQDFSRYE 300
```

```
QY 295 YGPEIRTRERLCROVLMYHHQLQALDSKITEHNGPTLVSRLLIINLYDESATASTLVFVRRVGH 354
D 301 YGFELTRRLCRQILMFHRLQTLGSOAKGDDPEALVSRLLILDYDENAMVSTLVSVRRVGH 360
QY 355 EQDGNVVVLPPELAYAQDFSPRHHAHQPMVDVLANFNAIQWQLVLDLKGELPGLLYQDK 414
D 361 EDNNVTVALPPELAYAQDFEQATLWQSDVLANFNFTIQWQLLDLKGEGVPGILYQDR 420
QY 415 GAWYRSQRIGESDVAVTWEKMQPLSVIPSLQSNASLVDPINGDGLDWVITGPGLRGY 474
D 421 NGWYRSQRQAQGBEMNAVTVGKMQLLPITPAVDQNASLMDINGDGLDWVITGPGLRGY 480
QY 475 HSQPDGSGWTRFTPLNALPVEYTHPRAQLADLMGAGLSDLVLIQPKSVRLYANTRDGPAK 534
D 481 HSQPDGSGWTRFTPLHALPIEYSHPRQAQLADLMGAGLSDLVLIQPKSVRLYVNRNDRGTE 540
QY 535 GKDVQSGSDITLPGADPRKLVAFSDVLGSGQAHLVEVSATKVTWCNPNLGRGFGQBIT 594
D 541 GRDVQSGSDITLPGADARKLVAFSDVLGSGQAHLVEVSATVTCWPNLGHGFGQBIT 600
QY 595 LPGFSQPAFENPAQVYLADLGGSGPTDLIYVHTNRDLIFLNKSGNGFAEPVTLRPEGL 654
D 601 LPGFSQSAASFNPORVHLADLGGSPADLIYVHADRLDIFNKGNGFAEPVTLRPEGL 660
QY 714 RFDHTCQLOMADVOGLGVASLILSVPHMSPHHWRCDLTNMPKFWLLNENANNNGVHTLRY 714
D 661 RFDHTCQLOMADVOGLGVVSVILSVPHMAPHHWRCDLTNMPKFWLLNENANNNGVHTLRY 720
QY 715 RSSQFQWLDKAAALTTGQTPVCVLPPIHTLMOTETEDISGNKLVTLRYARGAMDGR 774
D 721 RSSQFQWLDKAAALTTGQTPVCVLPPIHTLMOTETEDISGNKLVTLRYARGAMDGR 780
QY 775 EREFRGFGYEQTDSHQAQGNAPERTPPALTKWYATGLPVIDNALSTYWR-DDQAFA 833
D 781 EREFRGFGYEQTDSHQAQGNAPERTPPALTKWYATGLPVIDNALSTYWR-DDQAFA 840
QY 834 GFSRPTTWQNDKQVLTPEDDNRYWNRALKQGLLSELYGLDDSTNKHVPTVTFE 893
D 841 GFTFRPTLWKGKQVLTPEDDHNLWNLNALKQPLSELYGLDGSAAQOIPYVTFESR 900
QY 894 SOVRRLQHTDSRYVPLWSSVSESNRYHYERTASDPQCSQNTLSSDRFGQPLKOLSVQYP 953
D 901 PQVRLQDQATVSPVLWASVSESNRYHYERTASDPQCSQNTLSSDRFGQPLKOLSVQYP 960
QY 954 RROQPAINLYPDTLPDKLLANSYDDQQRQLRLTYQQSSWHHLTNVTVLGLPSTRSDI 1013
D 961 RNNKPTTNPYDTPDLTLPASSYDDQQLRLTLCQSSWHHLI GNELEVLGLPSTRSDA 1020
QY 1014 FTYGAENVPAQGLNLELLSDKSLIADDPKREYLGQQTAVTDGQNTTPTLQTPTRQALIA 1073
D 1021 FTYDAKQVPVVDGLNLETLCAENSLIADDPKREYLNQRTFTYTDGKNQTPPLKTPTRQALIA 1080
QY 1074 FTEITVFNOSTLSAPNGSI PSDKLSTTLLEQAGYQQTNYLFPRTGEDKVVVAHGHYDXT 1133
D 1081 FTEITAVTESLSPADGGITPDELPIQTQAGYQOEPYLPRTGENKVVVAHQYDXT 1140
QY 1134 AAQFWRPQKQNTQITGKITLIMDANYCVVQVTRDAAGLTASAKYDWRFLTPVQLTIND 1193
D 1141 EAQFWRPVPAQRNLSLTGKMTLKWDTYCVITQTDAGLTASAKYDWRFLTPVQLTIND 1200
QY 1194 NQHLITLDALGRPTLRFWGTENGKMTYSGPSEKASFPSPDVNAAIELKPLPVAQCV 1253
D 1201 NVHLITLDALGRPVTQRFWGTENGKMTYSGPSEKASFPSPDVNAAIELKPLPVAQCV 1260
QY 1254 YAPESWPEVLQKTFNRLAEODMQLYNARIIITEDGRICTLAYRWQSQAKIQLISLL 1313
D 1261 YAPESWPEVLQKTFNRLAEODMQLYNARIIITEDGRICTLAYRWQSQAKIQLISLL 1320
QY 1314 NNGPRLPHSITLTTDRYDHPPEQIQVQVFSFGFGLLQAAARHEAGMARQNEGSL 1373
D 1321 TNSICLPHNLTITTDYDRDSEQIQVQVFSFGFGLLQAAARHEAGMARQNEGSL 1380
```

```
QY 1374 IINVOHTENWAVTGRTEYDNKGQPIRTYQYFPLNDWRVYVNSDSARQEKAYADTHVDP 1433
D 1381 VTKVTKTIRWAVTGRTEYDNKGQPIRTYQYFPLNDWRVYVNSDSAR--KEAYADTHVDP 1438
QY 1434 IGREIKVITAKGWPRRTLTFTWFTVNEDENDTAAE 1468
D 1439 IGREIVITAKGWLRSQYFWFTVSEDENDTAAAD 1473

RESULT 5
QY 07N936 PHOLL PRELIMINARY; PRT; 1485 AA.
D 07N936 PHOLL PRELIMINARY; PRT; 1485 AA.
QY 01-MAR-2004 (TrEMBLrel. 26, Created)
D 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
QY 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
D 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
QY Name=tcaC; OrderedLocusNames=plu0515;
D Name=tcaC; OrderedLocusNames=plu0515;
QY Photorhabdus luminescens (subsp. laumondii).
D Photorhabdus luminescens (subsp. laumondii).
QY Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
D Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
QY NCBI_TaxID=141679;
D NCBI_TaxID=141679;
QY (1)
D (1)
QY NUCLEOTIDE SEQUENCE.
D NUCLEOTIDE SEQUENCE.
QY STEAIN-TT01;
D STEAIN-TT01;
QY MEDLINE=22957627; PubMed=14528314; DOI=10.1038/nbt886;
D MEDLINE=22957627; PubMed=14528314; DOI=10.1038/nbt886;
QY Duchaud E., Rusniok C., Frangeul L., Buchrieser C., Givaudan A.,
D Duchaud E., Rusniok C., Frangeul L., Buchrieser C., Givaudan A.,
QY Taourit S., Bocs S., Boursaux-Eude C., Chandler M., Charles J.-P.,
D Taourit S., Bocs S., Boursaux-Eude C., Chandler M., Charles J.-P.,
QY Dassa E., Derose R., Deruelle S., Freyassin G., Gaudriault S.,
D Dassa E., Derose R., Deruelle S., Freyassin G., Gaudriault S.,
QY Medigue C., Lanois A., Powell K., Signier P., Vincent R., Wingate V.,
D Medigue C., Lanois A., Powell K., Signier P., Vincent R., Wingate V.,
QY Zouine M., Glaser P., Boenare N., Danchin A., Kunat F.;
D Zouine M., Glaser P., Boenare N., Danchin A., Kunat F.;
QY "The genome sequence of the entomopathogenic bacterium Photorhabdus
D "The genome sequence of the entomopathogenic bacterium Photorhabdus
QY luminescens."
D luminescens."
QY Nat. Biotechnol. 21:1307-1313(2003).
D Nat. Biotechnol. 21:1307-1313(2003).
QY EMBL; BX571860; CAB12810.1; -; Genomic_DNA.
D EMBL; BX571860; CAB12810.1; -; Genomic_DNA.
QY Photolista; plu0515; -;
D Photolista; plu0515; -;
QY GO; GO:0005737; Cytoplasm; IEA.
D GO; GO:0005737; Cytoplasm; IEA.
QY GO; GO:0008305; C:integrin complex; IEA.
D GO; GO:0008305; C:integrin complex; IEA.
QY GO; GO:0007160; P:cell-matrix adhesion; IEA.
D GO; GO:0007160; P:cell-matrix adhesion; IEA.
QY InterPro; IPR000413; Integrin alpha.
D InterPro; IPR000413; Integrin alpha.
QY InterPro; IPR000408; Reg chr condens.
D InterPro; IPR000408; Reg chr condens.
QY InterPro; IPR003284; Sal_SpV.
D InterPro; IPR003284; Sal_SpV.
QY Pfam; PF01839; FG-GAP; 2.
D Pfam; PF01839; FG-GAP; 2.
QY Pfam; PF03534; SpvB; 1.
D Pfam; PF03534; SpvB; 1.
QY PRINTS; PR01341; SALSPVBPROT.
D PRINTS; PR01341; SALSPVBPROT.
QY PROSITE; PS00626; RCC1_2; UNKNOWN_1.
D PROSITE; PS00626; RCC1_2; UNKNOWN_1.
QY Complete proteome.
D Complete proteome.
QY SEQUENCE 1485 AA; 166091 MW; 7BB5EB1AD50B89A1 CRC64;
D SEQUENCE 1485 AA; 166091 MW; 7BB5EB1AD50B89A1 CRC64;

Query Match 58.6%; Score 4627.5; DB 2; Length 1485;
Best Local Similarity 58.4%; Pred. No. 3.7e-286;
Matches 869; Conservative 223; Mismatches 368; Indels 29; Gaps 11;

QY 1 MONSDQSFITELSLPKGGGATGNGEALTTPDGMALSLPLPISAGRGVAPFTLNYN 60
D 1 MONSDQSFITELSLPKGGGATGNGEALTTPDGMALSLPLPISAGRGVAPFTLNYN 60
QY 1 MODSPESVITLTLSPKGGGANGGALSAAGPDGMATLPLPLSTGRTAPGLSLYS 60
D 1 MODSPESVITLTLSPKGGGANGGALSAAGPDGMATLPLPLSTGRTAPGLSLYS 60
QY 61 SGAGNSPFLGWDGCVNVTIRRRTHFGVPHYDETDFTLFGPEGEVLWA-----DQPRDE 113
D 61 SGAGNSPFLGWDGCVNVTIRRRTHFGVPHYDETDFTLFGPEGEVLWA-----DQPRDE 113
QY 61 SSAGNGFPGIGWQGVWITISRRTHQHGIPQYGNDDTFLSPQGEVNVNIALNQGQPDIRQDV 120
D 61 SSAGNGFPGIGWQGVWITISRRTHQHGIPQYGNDDTFLSPQGEVNVNIALNQGQPDIRQDV 120
QY 114 STLAGINGLATFTVTGYRSLRLESHFSRLYEWQPKT--TGKTDFTWLIYSPDQVHLLGKSP 171
D 114 STLAGINGLATFTVTGYRSLRLESHFSRLYEWQPKT--TGKTDFTWLIYSPDQVHLLGKSP 171
QY 121 KTLQGVTLPISTVTRYQARQIVDFSRLEYWQSPASQGEGR--FWLISSPDQQLHILGKTA 179
D 121 KTLQGVTLPISTVTRYQARQIVDFSRLEYWQSPASQGEGR--FWLISSPDQQLHILGKTA 179
QY 172 QARISNSQTTQTAQWLLLEASVSRGQIYYOYRAEDDTGCEADEITHLQATQRYLHI 231
D 172 QARISNSQTTQTAQWLLLEASVSRGQIYYOYRAEDDTGCEADEITHLQATQRYLHI 231
QY 180 QACLANPQNDQIAQWLLLEETVPTGSHVSYQYRAEDETHCDNNEKTAHPNATAQRYLVQ 239
D 180 QACLANPQNDQIAQWLLLEETVPTGSHVSYQYRAEDETHCDNNEKTAHPNATAQRYLVQ 239
QY 232 VYGNRTASFTLPGDLSGASQADWFLVYVDFYDERSNNLKTTPAFST--TGSWLCRODRF 290
D 232 VYGNRTASFTLPGDLSGASQADWFLVYVDFYDERSNNLKTTPAFST--TGSWLCRODRF 290
QY 240 VYGNIRKQPSLFLVDNTPTPEEWLPHLFDHGERDTSLSHTVPKWDAGTAQWPRQDIF 299
D 240 VYGNIRKQPSLFLVDNTPTPEEWLPHLFDHGERDTSLSHTVPKWDAGTAQWPRQDIF 299
```

```
QY 291 SRYEYGEIRTRCLRQVLMYHHLQALDSDKITEHNGPTLVSRILNLYDESALASTLVFVR 350
Db 300 SRYEYGEVTRTRCQQLVLMYHHLQALDSDKITEHNGPTLVSRILNLYDESALASTLVFVR 359
QY 351 RVGHEQGNVVTLPPLLEYQDPSRPHAHWQPMVLAFNATQRLQVLDLKEGELPGLL 410
Db 360 QLSHEPDGSPITLPLLEAWQRENLKMPWTQWRFDAIDNFNSQORYQLVLDLKEGELPGLL 419
QY 411 YQDKGAWWYRSACRLGEIGSDAVTWKMKOPLSVIPSLQSNASLVLDINGDQLDMLVITGPG 470
Db 420 YQDKGAWWYRSACRLGEIGSDAVTWKMKOPLSVIPSLQSNASLVLDINGDQLDMLVITGPG 479
QY 471 LRGYHSQRPDGSWTRTFPLNALPVEYTHPRAQLADLMGAGLSDLVLIGPKSVRLYANTRD 530
Db 480 IRGYHSQRPDGSWTRTFPLNALPVEYTHPRAQLADLMGAGLSDLVLIGPKSVRLYANTRD 539
QY 531 GFAGKDVQVSGDITLVPVQADPRKLVAFSDVLGSGQAHVVEVSATKVCWPNLGRGRPG 590
Db 540 GMRKQDVQVSGDITLVPVQADPRKLVAFSDVLGSGQAHVVEVSATKVCWPNLGRGRPG 599
QY 591 QPITLPGFSOPATEFPAQVYLADLQSGGPTDILYVHTNRLDIFLANKSGNGFAEPVTLRF 650
Db 600 QPITLPGFSOPATEFPAQVYLADLQSGGPTDILYVHTNRLDIFLANKSGNGFAEPVTLRF 659
QY 651 PEGRLFDHTCOLQWADYQVGLGASLILSVPHMSPHHWRCDLTNNKMWLLNEMNNMNVH 710
Db 660 PEGVQDNTCOLQWADYQVGLGASLILSVPHMSPHHWRCDLTNNKMWLLNEMNNMNVH 719
QY 711 TLYRSSQWELDEKAALVTGQVPCVLPFPHTLWQTEDEISGNKULVTLRVARGA 770
Db 720 TLYRSSQWELDEKAALVTGQVPCVLPFPHTLWQTEDEISGNKULVTLRVARGA 779
QY 771 WDGREREFRGVVEQDTHQLAQGNAPERTPALTKNWTATGLPVDNALSTEYWR-DD 829
Db 780 WDGREREFRGVVEQDTHQLAQGNAPERTPALTKNWTATGLPVDNALSTEYWR-DD 839
QY 830 QAFAGFSPTTWO--DNKDVPITPDDNSRYNFNRALKQGLRSELYGLDDSTNKHVPY 887
Db 840 QAFAGFSPTTWO--DNKDVPITPDDNSRYNFNRALKQGLRSELYGLDDSTNKHVPY 898
QY 888 TVTEFSQVRLRLOHTSRYVPLWSSVVSRYHYRIASDPQCSQNTLSSDRFGPGLKQ 947
Db 899 TVTESRYQVRSIPVEKETELSAWVATENRSYHYRIISDPQLSQIRLQHDIFGQSLOS 958
QY 948 LSVQYPRQCPAINLPDITLPPDKLLANSYDQQRLELTYYQSSWHLLTNNYVRLGLPD 1007
Db 959 VDIAMPREKPAVNPYPITLPEALFDSYDDQQQLRLVRQKNSWHLLTDGENWRLGLPN 1018
QY 1008 STRSDIPTYGAEVNPAGLNLLELSDKNLSIADDPKREYLGQOKTAYTDGQNTPTLQPT 1067
Db 1019 AQRRDVTYDRTKIPTREGISLEVLKADGLLADEKAAYVIGQQQFTYTAGQSEVTLKPT 1078
QY 1068 RQALIAFTETTVNQSTLSAFNGSIPSDKLSLTLLEQAGYQOTNYLFPRTGEDKVVWAH 1127
Db 1079 LQALVAFQETAMWDDTSLQAYEGVIEBQGLNTTLTQAGYQVVARLENTGSESPVVARQ 1138
QY 1128 YTDYGTAAQFWRQKQNSQNTQITGKITLIWDANVCVVVOTRDAAGLTTSYAKYDRFLTPVQ 1187
Db 1139 YTDYRDATQFWRPQTRNSLLTGKTLTWTWTHCVIITQDAAGLTTSYAKYDRFLTPVQ 1198
QY 1188 LTDINDNQHILITDALCRPTTLRFWGTENGWMTGYSSEPKASFPSPDDVNAALTELKPLP 1247
Db 1199 LTDINDNQHILITDALCRPTTLRFWGTENGWMTGYSSEPKASFPSPDDVNAALTELKPLP 1255
QY 1248 VAQCQVYAPESWMPVLSCQTFNRL---AEQDWOKLYNARIITEDGRICTLAYRRWQSQK 1304
Db 1256 VAQCQVYAPESWMPVLSCQTFNRL---AEQDWOKLYNARIITEDGRICTLAYRRWQSQK 1315
QY 1305 AIPOLILSILNNGRLPHSLTLTDRYDHDPEQIRQOVVPSDGFGLLQAAARHEAGMA 1364
Db 1316 LTQLISLASISIPRLPHVIGITTDYDNDPQQHQQTYSFSDGFGHLLQSSVVRHPGDA 1375
QY 1365 RQRNEDGSLIINVQ-----HTENRWAVTGRTEYDNDKQPIRTYQPYFLNDRWYVNSDSA 1418
```

```
Db 1376 WORKEDGGLVNDANGALVSAATDTRWAVSGRTEYDDKGQVRYQPYFLNNRYVSDDSA 1435
QY 1419 ROEKRAYADTHVVDPTGREIKVITAKGWFRRTLTFTPWFTVNEDENDTAA 1467
Db 1436 RD--DLFADTHYDPLGREIKVITAKKYUREKQYTPFWFIVSEDENDTAS 1482

RESULT 6
ID O85153_PHOJU PRELIMINARY; PRT; 1485 AA.
AC O85153;
DT 01-NOV-1998 (TREMELrel. 08, Created)
DT 01-NOV-1998 (TREMELrel. 08, Last sequence update)
DT 01-FEB-2005 (TREMELrel. 29, Last annotation update)
DE Insecticidal toxin complex protein TcaC (Toxin complex protein).
GN NamestcaC;
OS Photorhabdus luminescens (Xenorhabdus luminescens).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Photorhabdus.
OX NCBI_TaxID=29488;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=W-14;
RX MEDLINE=98306208; PubMed=9641921; DOI=10.1126/science.280.5372.2129;
RA Bowen D.J., Rocheleau T.A., Blackburn M., Andreev O.,
RA Bhartiya R., ffrench-Constant R.H.;
RT "Insecticidal toxins from the bacterium Photorhabdus luminescens.";
RL Science 280:2129-2132 (1998).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=W-14;
RX Bowen D.J., Rocheleau T.A., Blackburn M., Andreev O.,
RA ffrench-Constant R.H.;
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=W14;
RX MEDLINE=21185117; PubMed=11286884; DOI=10.1016/S0966-842X(01)01978-3;
RA Waterfield N.R., Bowen D.J., Fetherston J.D., Perry R.D.,
RA ffrench-Constant R.H.;
RT "The tc genes of Photorhabdus: a growing family.";
RL Trends Microbiol. 9:185-191 (2001).
DR EMBL; AF046867; AAC38625.1; -; Genomic DNA.
DR EMBL; AF346497; AAL18451.1; -; Genomic DNA.
DR GO; GO:0005737; Cytoplasm; IEA.
DR GO; GO:0008305; C:integrin complex; IEA.
DR InterPro; IPR000413; Integrin alpha.
DR InterPro; IPR000408; Reg chr Condens.
DR Pfam; PF01839; FG-GAP; 2.
DR Pfam; PF03534; SpvB; 1.
DR PRINTS; PR01341; SALSPVBPROT.
DR PROSITE; PS00626; RCCL 2; UNKNOWN 1.
SQ SEQUENCE 1485 AA; 16616 MW; 74228E27A9921EBD CRC64;

Query Match 58.2%; Score 4595.5; DB 2; Length 1485;
Best Local Similarity 57.8%; Pred. No. 4; le-284;
Matches 861; Conservative 224; Mismatches 375; Indels 29; Gaps 11;

QY 1 MONSDQSFITSLPLKGGGALTQMGALTPGPDGMAALSPLPIAGRGYAPAFATLYN 60
Db 1 MQDSPEVSITSLPLKGGGALNGALNAGPDGMAALSPLPIAGRGYAPAFATLYN 60
QY 61 SGAGNSPFGLGWDCNVTIRRTTHFGVPHYDETDTFLGPEGEVLVVA-----DQPRDE 113
Db 61 NSAGNGPFGIGWCGVMSISRRTHQHGIPQXGNDTFLSPQGEVMNIALNDGQPDIRQDV 120
QY 114 STIQNLGNATFTVTGYRSLESHFSRLFWQPKT--TGKTDFWLIYSPDQVHLGKSP 171
Db 121 KTLQGVTLPISTVTRTQARQILDFSKIEYQWQASQGRGA-FWLISPPDQLHLGKTA 179
```

Qy 172 QARLSNPSTOTTAQWLLLEASVSRGEQIYYOYRAEDDTGCEADBIHTHLOAQORVLIHI 231  
Db 180 QACLANPONDQOIAQWLLLEETVTPAGEHVSYOYRAEDDEAHCDNEKTAHPNVTAQRYLVQ 239  
Qy 232 VYGNRTASETLPGIDGSAPOAWLFLVLPDYDYSRNNLKTTPAFST-TGSMCLCRODRF 290  
Db 240 VNYGNIRPOASLFLVLDNAPPAPEEWLPHLVFDHGERDTSLHTVPTWDAGTAQMSVRPDI 299  
Qy 291 SRYEGFEIRTRRLCRQVLYHHQLADSKITEINGPTLVSRLLINVDSEIAISTLVFVR 350  
Db 300 SRYEGFEVTRRLCQVLMFHTALMAGEASTNDAPELVRLILEYDKNASVTTLITIR 359  
Qy 351 RVGHEODGNVTLPLELAYODSPRHHAHQWMDVLANFNAIORWOLVDLKGSLPCLL 410  
Db 360 QLSHESDGSPTQPLELAWORFLEKMTWQRFALONFNSQORYQLVLRGSLPGML 419  
Qy 411 YQDKGMYRQAORLGEIGSDAVTWKMQPLSVIPSLQSNASLVMDINGDQDQWVITGPG 470  
Db 420 YQDRGMYKAPQOEDGDSNAVYDKIAPLPTLPNLDNASLMDINGDQDQWVITGAS 479  
Qy 471 LRGHYSQRPDGSWTRFTPLNALPVEYTHPRAQLADLMGAGLSDLVLIGPKSVRLYANTRD 530  
Db 480 IRGHYSQRPDGSWTRFTPLNALPVEYTHPRAQLADLMGAGLSDLVLIGPKSVRLYANTRD 539  
Qy 531 GFAKGKVVQSGDITLVPUGADPKLVAFSDVLGSGQAHLYEVSATKVTCPNLRGRFG 590  
Db 540 GWRKGEDVPQGTGTLPTVTGTDARKLVAFSDVLGSGQAHLYEVSATKVTCPNLRGRFG 599  
Qy 591 QPITLPGFSQATBNPAQVTLADLDSGPTDLYVHTNRLDIPLNKGNGFAPSVTLRF 650  
Db 600 QPITLPGFSQATBNPAQVTLADLDSGPTDLYVHTNRLDIPLNKGNGFAPSVTLRF 659  
Qy 651 PEGRLFDHTCOLQWADVQGLVASLILSVPRMSPHHRCDITNNKPMLLNMMNNMGVHH 710  
Db 660 PEGVQFQNTCOLQWADVQGLVASLILSVPRMSPHHRCDITNNKPMLLNMMNNMGVHH 719  
Qy 711 TLRYSRSQFWLDEKAAALTTQGPVCLPPIHTLWCTETEDISGNKLVTLRYARGA 770  
Db 720 TLRYSRSQFWLDEKAAALTTQGPVCLPPIHTLWCTETEDISGNKLVTLRYARGA 779  
Qy 771 WDGERFRGYYEQTDHSHLOACNAPERTPPALTKWYATGLPVDNALSTBYWR-DD 829  
Db 780 WDGERFRGYYEQTDHSHLOACNAPERTPPALTKWYATGLPVDNALSTBYWR-DD 839  
Qy 830 QAFAGSPRFTWQ--DNKDYPLTPEDDNRYSYFNWALKGQLLSLYGLDDSTNKKVY 887  
Db 840 QAFAGSPRFTWQ--DNKDYPLTPEDDNRYSYFNWALKGQLLSLYGLDDSTNKKVY 898  
Qy 888 TVTFERSQVRLOHTDSRYFVLSVSVSRNHYHTASDQCQSONITLSDRFGQPLKQ 947  
Db 899 TVTFERSQVRLOHTDSRYFVLSVSVSRNHYHTASDQCQSONITLSDRFGQPLKQ 958  
Qy 948 LSVQYPRQOAINLYPDTLPDKLLANSYDQOQLRLTYOQSSWHHLTNNTVRVGLPDP 1007  
Db 959 VDIAPRKEKAVNPYPTLPETLPDSYDQOQLRLTYOQSSWHHLTNNTVRVGLPDP 1018  
Qy 1008 STRSDIFTYGAENYPAGNLIELLSKNSLIADKPREYLGQOKTAYTDGQNTTFLPTPT 1067  
Db 1019 AQRDVTYDRSKPTGEGISLEILLKODGLLADEKAAVYLGQOKTAYTDGQNTTFLPTPT 1078  
Qy 1068 QOALIAFTETTVFNQSTLSAFNGSIPSKLSTTLQAGYQNTNLPFRGDKVWVAHHG 1127  
Db 1079 LQALVAFQETAMDDTSLQAYEGVIEQELNLTALTOAGYQOQVARENTRSESPVAAARQ 1138  
Qy 1128 YTDYGTAAQWRPQKQNTLTGKTLTLDWANCYVVTQTRDAAGLTTSKAYDWEPLTPVQ 1187  
Db 1139 YTDYGTAAQWRPQKQNTLTGKTLTLDWANCYVVTQTRDAAGLTTSKAYDWEPLTPVQ 1198  
Qy 1188 LTDINDNOHLITLDALGRPITLRFWGTENGQWTVSSPEKASFPSPDVNAAILKPKPLP 1247  
Db 1199 LTDINDNOHLITLDALGRPITLRFWGTENGQWTVSSPEKASFPSPDVNAAILKPKPLP 1255  
Qy 1248 VAQCOVAPESMVPVLSOKTNRL---ABQWQKLNYNARIITEDGRICITLAYRNRWQSQK 1304

Db 1256 VAQCLVAVDSWMPSSLQSLQSQSEAEALWAQLRAAHMITEGKVCALSGKRGTSQHN 1315  
Qy 1305 AIPQLISLNGPRLPPLPHSLTLTTDRYDHPDQEQIRQOVVFSQDGFRLQLQAAARHEAGMA 1364  
Db 1316 LTIQLISLNASIPRLPHVLTITDRYDSDPQQHQQTQVFSQDGFRLQLQAAARHEAGMA 1375  
Qy 1365 RQNEDESLINVO-----HTENRVAOTGRTYDNGKQPIRTYQYFPLNDWRVYVND 1418  
Db 1376 WORKEDGLVVDANGVLVSAFTDTRWAVSGRTYDNGKQPIRTYQYFPLNDWRVYVND 1435  
Qy 1419 ROKEKAVADTHVDPDPIGRIKVTAKGFRRTLTFTWFTVNEDENDTAA 1467  
Db 1436 RD--DLFADTHLDPDPIGRIKVTAKGFRRTLTFTWFTVNEDENDTAA 1482  
RESULT 7  
Q6XP56 YERFR  
ID Q6XP56\_YERFR PRELIMINARY; PRT; 1433 AA.  
AC Q6XP56;  
DT 05-JUL-2004 (TrEMBLrel. 27, Created)  
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)  
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)  
DE TcYF2  
OS Yersinia frederiksenii.  
CG Plasmid pMY1.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Yersinia.  
OX NCBI\_TaxID=29484;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=49;  
RA Dodd S.J., Clare T.R., O'Callaghan M., Ronson C.W.;  
RT "Yersinia enterocolitica contains plasmid-encoded genes that show  
RL similarity to the tc family of insecticidal toxins."  
RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AY220492; AAF57764.1; -; Genomic\_DNA.  
DR GO; GO:0005737; C:cytoplasm; IEA.  
DR GO; GO:0008305; C:integrin complex; IEA.  
DR GO; GO:0007160; P:cell-matrix adhesion; IEA.  
DR InterPro; IPR000413; Integrin alpha.  
DR InterPro; IPR003284; Sal\_SpVb.  
DR InterPro; IPR006530; YD.  
DR Pfam; PF01839; FG-GAP; 2.  
DR Pfam; PF03534; SpvB; 1.  
DR PRINTS; PR01341; SALSPVBPROT.  
DR TIGRFAMs; TIGR01643; YD\_repeat\_2x; 1.  
KW Plasmid.  
SQ SEQUENCE 1433 AA; 157723 MW; 15FDEFA15118A4B CRC64;  
Query Match 50.4%; Score 3984.5; DB 2; Length 1433;  
Best Local Similarity 52.8%; Pred. No. 4.3e-245;  
Matches 786; Conservative 187; Mismatches 432; Indels 85; Gaps 15;  
Qy 1 MONSDPSITELSLPKGGGAIITGMEALTPGDPGMAALSILPLIPISAGRYAPAFITLYN 60  
Db 1 MONRQDMAITTPSLPSSGGAVTGLKGDIAAAGPGAATLSIPLVSPGRTPTGTLYNH 60  
Qy 61 SGAGNSPFLGMDCNVMTIRRTHFVPHYDDETFGLGPEGEVLVA-----DQPD 114  
Db 61 SRAGNTFGTGLSGGPAVQVRRTRSGAPAYDETFGPDGGEVLVPAITAGTPTETQAT 120  
Qy 115 TLQINLGAFTVTGVRSLRSHFSRLRYVQPKTKTGTDFWLIYSPGQVHLLGKSPQAR 174  
Db 121 TLGIFSGNFSQVYRSRTGSLRSLRWVPADEAETDFWLYTPDQVALLGNGAAR 180  
Qy 175 ISNPSQTTQAQMLLEASVSRGEIYYQYRABDDTCCRADEITHHLOAQRYLHVY 234  
Db 181 ISNPARPTQAVLMSSVSLTGEQMYQYRABDDTCCDEMRDAHPQAGRYLVAVY 240  
Qy 235 GNRTASETLPLDGSAPSQADWLFYLYFDYGERNNLKTTPAFSTG--WLCHODRFSR 292  
Db 241 GNROARALPAL--MSTPSMDSWLPVQVFDYGERSTALSEPPAWQTPGSGEWRRCQDCFSE 299





Qy	61	SGAGNSPGLGWCDCNVMTIRBTHFGVPHYDETDTFLGPCEGVLWA-----DQPRDES	114
Db	61	SRSGNGPFGIGWGIGGAAVQRTNRNGAPTYDDTDEFGPDGEVLVPALTAAGTQEAHQAT	120
Qy	115	TLOGINLGATVTVTGYRSRLSHESHPRSLEYMQPKTTGKTDFWLIIYSPOGVHLLGKSPQAR	174
Db	121	SLGGINPGGSFNVQYRSRTSGLSRLERWLPADETETEFMWLYTPDGOVALLGNRAQAR	180
Qy	175	ISNPSQTTQRAQWLLLEASVSRGBOIYYQYRAEDDTCGEADEITHLQATQRYLHVYY	234
Db	181	ISNPTAPTQTAVWLMESSVSLTGQMYQYRAEDDDGCDEAERDAHPQAGAQRYPVAVWY	240
Qy	235	GNRYSASETLCGLDGSAPSQADWLFLVFDYGERSNLKTTPAFSTGCS--WLCRODRFSR	292
Db	241	GNRQARITLPAL-VSTTFSMDSWLFLVFDYGRSSVLSEAPAWQTPSGGEWLCKQDCFSG	299
Qy	293	YEYGFETRRLCRQVLMYHHLOALDSKITEHNGPTLVSRILILNYDESATISTASTLVFVRRY	352
Db	300	YEFGNLRTRLCRQVLMFHYLVGLAGSSGANDAPALISRLLLDYRESPSLLENHVQV	359
Qy	353	GHEODGNVVTLPPELAYQDPSPRHHAHWQPMVLAMFNALQRMQLVDLKEGELPGLLYQ	412
Db	360	AYESDGTSCALPALALGQMTPTPTLSAWQTRDDMGKLSLLQPYQLVDLNGEVGVGLIYQ	419
Qy	413	DKGAWYRSRQRLCEIGSDAVTWKMQPLSVIPSIQSNASLVDINGDOLDWMTTGBCLR	472
Db	420	DSGAWWTREPVQRGGDDPDVWGAQAALPTMPALHNSGILLADLNGDRLEWVVTARGVA	479
Qy	473	GYHSQRPDGSWTRFTPLNALPVEYTHPRAQLADLWMGAGLSDLVLIGPKSVRLYANTDRGF	532
Db	480	GMYDRTPGRDWLHFTPLSALPVEYAHPKAVLADILGAGLITDMVLIGPRSVELYSKNDGW	539
Qy	533	AKGXDVVQSGDITILPVGADPRKLVASPDVLGSGQAHVVEVSATKVTCPWNLGHRGRCQP	592
Db	540	NKGSTVQOTERLTIIPVFGVDPRTLVAFSDMAGSQOHLTEVRANGVRYVWPNLGHGRGCP	599
Qy	593	ITLPGFSOPATEFNPAPVYLADLDSGPTDLIIYVHTNRDLITFLNKSNGNGPAEPTVTLRPE	652
Db	600	VNIIFGFSQSVTTFNPDQILLADTQSGTDTLLIYAMSRLVLIYFNQSGNYFAPHTLLLPK	659
Qy	653	GLRFDHTCQLQADVQGLGVASLILSVPHMSPHWRCDLTNMKPWLLNMNNMNGVHRTL	712
Db	660	GVRDRTCSLQVADIQGLGVPSLLITPVHAPHHVWCHLSADKPWLLNGMNNMNGARHAL	719
Qy	713	RYRSSQFWDLDEKAALTTGQTPCYLPPPTHTLWQTEDEISGNKLVITLRYARGAWD	772
Db	720	HYRSSVQFWDLDEKAEALAGSSPACYLPFTLHTLWRSVVQDEITGNRLSDVLRYHGVM	779
Qy	773	GRERFRFGVVEQTDGSHOLA-CQNPABERTPPALTKWYATGLPVIDNALSTEYWRDD-Q	830
Db	780	GOERFRFGFVEIRDDTTLASQGTATLSNPSSVSRNWYATGVPAVDERLPEITWQNDAA	839
Qy	831	AFAGFSRFTTWQNDKVPLTPEDDNSRYWFNRALKQQLRASELYGLDDSNKCHVPYTVT	890
Db	840	AFADFATRFTVSGSEDEQTYTP-DDSKTFWLQRLKGLLRSELVYADGSSQADIPYSVT	898
Qy	891	EFRSQVRRLQHTDSRYPVJLSSVBSRNYHYERTASDPQCSQNTILSDRFGQPLQSLVS	950
Db	899	ESRPQV-RLVEANGDYPVWPMGAEBSVTSVYTHNDPQCCQQAVLLSDEYGFRLQVSV	957
Qy	951	QYPRRQCPAINLYPDTLLPKLLANSYDDQQRQLRLTYQSSSWHHLTNNTVR--VLGLPDS	1008
Db	958	NYPRRPPSADNPYPASILPATILFANSYDEQQOILLGLIQSSSAHHLVLSLSEGHWLLGLAEA	1017
Qy	1009	TRSDIFTYGAEVNPAGGLNLELLSDKNSLADDRPREVLGQOKTAYTDGQWTTPLQTPTR	1068
Db	1018	SRDDVFTYSADNVEPGLLTLEHLLAPESLVSDSQVGTLAGQQQWYLDSDQVATVAAPPL	1077
Qy	1069	QALLTAFTTETTVFNOSTLSAFNGSTPSDKLSTTLTBOAGVQQTNYLP--RTGEDKVVWAHH	1126
Db	1078	PPKVAFTETALDSEGMVSSLAAYTVDEH----LEQAGVRSQGYLFPFGREAEQALWQCQC	1133

Qy	1127	GYTDYGTAAQFWPKQNSNTQLTKGLTLIWDANYCVVQVTRDAAGLTTSKAYDWRFLTPV	1187
Db	1134	GYVTYGAAGHFWLPLSPRSDNMLGTPVTTRDADYDCVITQWQDAAGIVTTADYDWRFLTPV	1193
Qy	1187	QLTDINDNQHLITLDALGRITITLRFWGTENGKMTGYSSPEKAFSPSPDVNAATELKKPL	1246
Db	1194	RVTDPNDNLQSVTLTDALGRVTTLRFWGTENGITATGYS--ATLSVPDGAAALALATPL	1250
Qy	1247	PVAQCYVAPESWMPVLVSQKTFIRLAEQDWQKLYNARIITTEDGRICTLAYRRVWQSOKAI	1306
Db	1251	PVAQCLVYVYDSW-----GDDNNEK-----	1270
Qy	1307	PLQISLILNGLRPLPHSLTLTTDRYDHPDQQLRQQVVSFSDGFGRLIQAARAEAGMARQ	1366
Db	1271	-----NPPHVVLATDRYDSDTGQVVRQVTFSDGFGRELOSATRQAEAGNAWQ	1318
Qy	1367	RNEDGSLI-----INVQHTENRWAVTGTEDYNDKQGPTRTTPYPLNDWRVYVNSDSAR	1419
Db	1319	RGRDGLVITASDGLPVTVTA-TNFRWAVTGRAEYDNKGLPVRVTPYQVFLDSWQVYVSDSAR	1377
Qy	1420	QEKAYADTHVYDPIGRIEIKVITAKGWFRTLTFTPWFTVNEDENDT	1465
Db	1378	Q--DLYADTHFYDPTAREWQVITAKGERRQVLYTPMPVNSEDENDT	1421

RESULT 9

Q6QDZ4\_YERPS

ID Q6QDZ4\_YERPS PRELIMINARY; PRT; 1489 AA.

AC Q6QDZ4; Q66505;

DT 05-JUN-2004 (TrEMBLrel. 27, Created)

DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)

DT 10-MAY-2005 (TrEMBLrel. 30, Last annotation update)

DE Y0185-like protein (Insecticidal toxin complex)

GN Name=tCaC; OrderedLocusNames=YPTB3553; ORFNames=YPP3678;

OS Yersinia pseudotuberculosis.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

OC Enterobacteriaceae; Yersinia.

OX NCBI\_TaxID=633;

RP NUCLEOTIDE SEQUENCE.

RP STRAIN=IP32953;

RA Bernardes Pinheiro V.B., Ellar D.J.;

RL Submitted (MAY-2004) to the EMBL/GenBank/DBJ databases.

RN [2]

RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].

RC STRAIN=IP32953 / Serotype 1;

RC PubMed=15359858; DOI=10.1073/pnas.0404012101;

RA Chain P.S.G., Carniel E., Larimer F.W., Lamerdin J., Stoutland P.O.,

RA Regala W.M., Georgescu A.M., Vergez L.M., Land M.L., Motin V.L.,

RA Brubaker R.R., Fowler J., Hinnebusch J., Marceau M., Medigue C.,

RA Simonet M., Chenal-Francoisgue V., Souza B., Dacheux D., Elliott J.M.,

RA Derbise A., Hauser L.J., Garcia E.;

RT "Insights into the evolution of Yersinia pestis through whole-genome

RT comparison with Yersinia pseudotuberculosis".

RT Proc. Natl. Acad. Sci. U.S.A. 101:13826-13831 (2004).

DR EMBL; AY345599; AAS66065.1; -; Genomic DNA.

DR EMBL; BX336398; CAH22791.1; -; Genomic DNA.

DR GO; GO:0005737; C:cytoplasm; IEA.

DR GO; GO:0008305; C:integrin complex; IEA.

DR GO; GO:0007160; P:cell-matrix adhesion; IEA.

DR InterPro; IPR004413; Integrin\_alpha.

DR InterPro; IPR004048; Reg\_chr\_condens.

DR Pfam; PF01839; PG-GAP; 1.

DR Pfam; PF03534; SpvB; 1.

DR PRINTS; PR01341; SALSPVBPROT.

DR PROSITE; PS00626; KC1\_2; UNKNOWN\_1.

DR Complete proteome.

QW SEQUENCE 1489 AA; 167592 MW; CEA5B7C0A04B7B44 CRC64;









Query Match		42.4%;	Score 3350.5;	DB 2;	Length 1401;	
Best Local Similarity		48.6%;	Pred. No. 1.3e-204;			
Matches 685;		Conservative 197;	Mismatches 474;	Indels 53;	Gaps 21;	
QY	1	MONSQPSITELSLPKGGGAIOMGEALITGPDGMAALSLPLPISAGRYAPAFITLNYN	60			
DB	1	MQGSTPLKLEIPSLPGGGSLKMGGEALNAVAGGASPSLPISVGRGLVPVLSLNY	60			
QY	61	SGAGNSPFLGWCNWTIRRTTHFGVPHYDETDITFLGPEEVL-VVAD---QP--RDES	114			
DB	61	STAGNSFGMGWCGVGFISLRKAGVPHYTGQDEVLGPDGEVLSIVPDSQGGPEQRTAT	120			
QY	115	TLOGINLGAFTVTGYSRSLRSHFSRLEYWQP--KTTGKTDFWLIYSPDCOVHLLGKSPQ	172			
DB	121	SLIGTVLTQHTVTRVQSRVAEKIVLEHWQPOORREETSFWLFTADGLVHLFGKHH	180			
QY	173	ARISNPSQTTQATOWLLEASVSRGEQIYYQYRAEDDTGCEADEITHH--LQATAQRYLH	230			
DB	181	ARIADPODETRIARLWMEETVTHGEHIYYHYRAEDDLDCEHELAQHSVLRPTLSWQ--	238			
QY	231	IYVYGNRTASSETPLGDCSAPSQADWLFLYVFDYGGERSNMLKTPPAPSTTGS-----	282			
DB	239	-VOYGNTQPTAFVAKSGIPVDNDWLFHLVDPYGERLSLNSVPFNVSENNVSNVS	297			
QY	283	--WLCRODRFSRYEYGEIRTRRLCROVMYHLOALDSKITEHNGPTLVLSRLILNYDES	340			
DB	298	EKWRCPDSFSRYEYGEIRTRRLCROVMYHLOALDSKITEHNGPTLVLSRLILNYDES	357			
QY	341	AIASLTIVFVRVRRHQEGDGNVVTLPPLLEAYQDFSPRHHAHWQMDVLNANAIQRMQLVD	400			
DB	358	NKVSLLQTARSLAHEITDGTVPVMSPLEMDYQRVNHGVNLWQSPQLEKNTLQPYQLVD	417			
QY	401	LKGEPLGLLYQD-KGAWTRSAQR-LGETGSDAVTWKMQPLSVIPSLQSNASLVINDG	458			
DB	418	LYGEGIPALLYQDTQKAWTRAPVRDITAEGTNAVTYEEAKPLPHIPAQOESAMLLDING	477			
QY	459	DGOLDWITGPGRLGYHSQRPDGSWTRFTPLNALPVEYTHPRAQLADLMGAGLSDLVLIG	518			
DB	478	DGRLDWITASGLNGYHNGPEGBWTFPIPLSAVPMFYFHPQAKLADIDAGLPDLALIG	537			
QY	519	PKSVRYLVANTRDGFPAKGVVQSGDITLPVPGADPRKLVAFSDVLGSGQAHLEVSATKY	578			
DB	538	PNSVRVWNNPAGWDRAQDVTHLSNKLPLVPGKHKRLHVAFSDMTGSQSHLVEVTANSV	597			
QY	579	TCWPNLGRGFGQBITLPGSQPATEFNPAQVILADLDGSGPTLLIYVHTNRDLIFLKS	638			
DB	598	RYWPNLGHGKFGELMITGFQITGKRLTPTDCIWTYTTMAQAPDPFIYARNTYLEYANES	657			
QY	639	GNGFAEPVTLRFPBGLRFDHTCOLQADVOGLGVASLILSVPHMSPHHWRCDLTNMKPWL	698			
DB	658	GNSHAEFORLDLPGVAFRFDTCRIQIADTQGLGTASLIILTPHMKVQHWRLDMTIFPWL	717			
QY	699	LNEMNNMGVHHTLYRSSQFWLDEKAAALTTGQTPVCYLPFPFIHTLWQTEDEISGN	758			
DB	718	LNAVNNMGTEITLYYSSAQFWLDEKQASQESGWTVVSYLPPFVHVMRTEVLDEISGN	777			
QY	759	KLVTTLRYAGAMDGRERFRFGYVQEQSHQLA---QGNAPERTPPALTQWYATGLP	815			
DB	778	RLTSHYVSHGAWDGLRERFRGFRVQTQDIDRSASATQGTTHAEPAPSRPTVWYGTGVR	837			
QY	816	VIDNALSTYWR-DDQAFAGSPREFTTQDNK--DVPLTTPEDDNRVYFNRAKGQLRS	872			
DB	838	EVDILLPTETWGGQQQAAPHITRPTFRYDEKSGDGMVTP-SEGEFWLHRAKGQLRS	896			
QY	873	ELYGLDSTNKHVYTVTEFRSQVRLRQHTDSRYFVLWSSVYSESNHYHRIASDPQCSQ	932			
DB	897	ELYGDDSDILAGTFYSVDESRTQVRLPLVMVSDVPAVLVSVAESQRYERYEVAITDPCSQ	956			
QY	933	NITLSSDRFQOPLKQLSVQVPRROQPAINLYPDTLPKLLANSYDDOQORLRLTYQSSW	992			
DB	957	KIVLKSALGPPQDNLEIAYSRRPQPSFPYDPTLPTLTSSPDEQMFRLRTRQPSY	1016			

Query Match  
Best Local Similarity 36.1%;  
Matches 544; Conservative 225; Mismatches 595; Indels 143; Gaps 37;

Score 2277.5; DB 2; Length 1447;  
Pred. No. 5.2e-136;  
Matches 544; Conservative 225; Mismatches 595; Indels 143; Gaps 37;





8Q	SEQUENCE	1446 AA; 162070 NW; 8DBAAD65DCC8949 CRC64;
	Query Match	28.8%; Score 2274; DB 2; Length 1446;
	Best Local Similarity	35.7%; Pred. No. 8.7e-136;
	Matches 550; Conservative 220; Mismatches 562; Indels 188; Gaps 39;	
Qy	2	QNSQDFITSLSPKGGGAIITGMEALTPGPGDMAALSPLPISAGRGYAPAFITLNYNS 61
Db	13	QSLQPAVATPL-LPKGGGAIQSIGKMGWSVGTSGAASLEVALPISPRGYPALSLYSQS 71
Qy	62	GAGNSPGLGNDGNMTIRRTTHGPHYDETTFFLGPEGEVLVADQPRDE----- 113
Db	72	TSNGVFLGWLNTSKVARASRGVPTYANDDLIFGPGGDVCL---PERDDSGALVSTQ 128
Qy	114	-STLQGINLCAFTVTCYRSLRSHESRLEVMQPKTKGTDWMLYSPDGOVHLGKSPQ 172
Db	129	VSRYNHDLNATVQVRYFRSEGAFLRIEHW-VDIADPGFWLHGADGSLNLYGRTS 187
Qy	173	ARISNPQTTQTQWLLLEASVSRGEQIYYQYRAEDDTGCEADBITHLQATQRYLHV 232
Db	188	SRIADPADMRVAEWLLDESMNAGEHILYEYKPEDHQGLAED---HPRNFRAQRYLSRV 244
Qy	233	YYGNRTASETL-----PGLDGSAPSQADWLFYLVFDYGERSNLKTPPAFSTTGSMLCRQ 287
Db	245	RYGNAKAHYYVYLWQEDSLDGLL-----WHFDLIFDYDQDTRSEPPPEYDEQFTWPRS 299
Qy	288	DRPSRYEYGEIRTRRCROVLYVHLQALDLSKITEHNGPTLVSRLLTYNDESAIATLV 347
Db	300	DPHSSFAYGPELGNLRCROVLMPHH---FPNELGE--APLLTRRLLEHLQTTLGYNLL 354
Qy	348	FVRVVGHEQ--DG-----NVVTLPLELAYQDFSPRHHAHQPMDFLANFNAIORWQLVDL 401
Db	355	---SAASHQAWDGTDRWRVQQPPVQYTDPS--LESGIYTPLEPMAGLNDGQQYQLVDL 410
Qy	402	KBGLPGLLYODKGAWYRQAQRIGEIGSDAVTWKMQPISVIPSLQS-----NASLVGIN 457
Db	411	YGDGLPGILYRDDKAMLYREPVRTTGCADAVAYGACQPLPRIPTADSATPVRTLTDLT 470
Qy	458	GDGOLDWITGCLRGVHSQRPDGSTRPTPLNALPVEYTHPRAQLADLACAGLSLAVLI 517
Db	471	GDGRLDWVVAQPGMAGPFTLNPDRSWSNYATFSAPFAEFPHQPGQMDLVGDGLSLALI 530
Qy	518	GPKSRLVYANTR--DGFAGKDVVQSGDITLVPFGADPRKLVAFSDVLGSGQAHLEVSAT 576
Db	531	GPRSVRLYANRRADGFAAATDIPHDED-RUPLLSDSSTELVAFSDLLGTGQQLIRIRHN 589
Qy	577	KVTCWPNLGRGRFCQP---ITLPFGSQPATEFNPQVYLADLDGSGPTDLIYVHNRKLDI 633
Db	590	EIRVWPNLGRGRFGKGLFATLP-YTYEA--FDSQVRLADLDGSGASDVLYLQADGFQV 646
Qy	634	FLNKSNGPAPVPTLRPEGLRPDHTCOLQWADVQGLGVASLIILSVPHMSPHWH----RC 689
Db	647	FMNRGGNGLAAPQORWPEGVRYDRFCQSAVDLLGLGFSLSLVTVPHMAPLHWSLYAA 706
Qy	690	DLTN--MKPMLLNMNNMNVHHTLRYRSSSQFWMDEKAAALTTGQTPVCVLPPIHTLW 747
Db	707	DRAGSVKPYLLKASDNNLGNAGEVYSRSAQEWLDEKNELRAAGSVAVSELPPVHVVV 766
Qy	748	QTEDEDISGNKLVTTILYARGAWDGRERFRGFGYVEQTDSSHOLAQGNAPERTPPALTK 807
Db	767	QTMQDKVTGNTLTQLFRYRQGFYDPRERFRGFGLLQTDTSQ--DQEDFTAPVLNK 825
Qy	808	NWYATG-----LPVIDNALSTEYWRDDQAFGFSRFTTWQ 843
Db	826	TWFTGTRYPARCTDHDRLDLARLPGEHVLRLSLRDLAATAQTEQ----- 867
Qy	844	DNKDVLPTPEDNDRYWFNRALKQLLRSELYGLDDSTNKGVPYTVTFRSQVRRLQHTD 903
Db	868	-----PITDADATLQEMARALSGSVLSEVFGLDASQRPVLYLSTRSVLYRQLQALS 922
Qy	904	SRYVLWSSV-----VESRNYHE--RIASDPOCSQNTILSSDRFQOPLKQLSVQYPRRQOP 958
Db	923	AHRP--YASMLPLSLLEVITYRYEAEELEDPMCEHSLNLAWDRYGSTLHSVSNYARRKCP 980

Qy	959	A-INLYPDTLPDKLLANSYDDQORURLTYQSSWHHLTNTNTRVRLGLPDPSTRSDIFTYG 1017
Db	981	GDAPPADPHQQQWEASHDDAQOQPYLNMHAEALYLDSPQSWRLGLPYRTRGDAMLIP 1040
Qy	1018	AENVPAAGLNLELLSKNSLIADDKPREYLGGQKTAFT---DGQNTTPIQTPTTROALIAF 1074
Db	1041	ASALTPAQISYEQFADPSGFFA-TLPRITLSLSVQRYIGCGDE-----ASFQALADA 1092
Qy	1075	TETTVNQSTLSAFNGSIPSDKLSLTLEQAGYQOQWYLPRTGEDKVMVAHHCYTDYGT 1134
Db	1093	VETAEELDHALSAYERVMDSVTLAEKLVIEIGYQQMPSFLPADSLN-LMSVKRGFATYAGQ 1151
Qy	1135	AOFWRPQKQSNTOCTGKITLIWDANCVVVVQTRDAAGLTTSKAYDWRFLTPVQLTDINDN 1194
Db	1152	EHEFHTQTPRFRSHGWSLVEYDAYHLFATRIDPADCVTTAEYDYRVLPQKRIIDPNQN 1211
Qy	1195	QHLITLDALGRPTTLRFMGTEGKMTGYSPEKASFPSPDVNAAIELKKPLPVAOCQVY 1254
Db	1212	RQADYDAFGRVWATSFYGTGELGEAVGPPPLNRAGHYWASAGEVALQ----- 1258
Qy	1255	APESWMPVLSQKTPNPLAESQDWQKLYNARIITEDGRICTLAYRRWTSQKAIIPQLISLLN 1314
Db	1259	-PE-----YALGRQ-----ASALYYDGN-----TVLGLVH 1382
Qy	1315	NGPRLPPHSLTLTTRDYDHDPEOQIRQQVVFSDGFRLLQAAARHEAGMARQNRNEDGSLI 1374
Db	1283	-----IPLATAVLVADRYPEDLDKQIRISMASIDGFGRTLTQKQVEDGDAYSVDWGNLE 1338
Qy	1375	I-----NVOHTENRWAVTQRTYDNKGQPIRTYQPYPLNDRMRYVNSDSARQEKEAYADTH 1429
Db	1339	LVDGPKIVHASPRWRISERVEYNNKGLAVRVVPYFANSHLYVNDASIRSQN--IVDKQ 1396
Qy	1430	VYDPIGREIKVITAKGWFRTLETPWFTVNEDENDTAAEV 1469
Db	1397	FYDPLGRPTTIITAKGMRRRQTYRVWYTTISEDENDTABEV 1436

Search completed: December 16, 2005, 14:41:34  
Job time : 281 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 16, 2005, 14:25:20 ; Search time 211 Seconds  
(without alignments)  
3069.404 Million cell updates/sec

Title: US-10-706-424-10  
Perfect score: 7901  
Sequence: 1 MQNSQDSITELSLPKGGGA.....WFTVNEDENDTAAEVKKVKM 1474

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_21:\*

- 1: Geneseqp1980s:\*
- 2: Geneseqp1990s:\*
- 3: Geneseqp2000s:\*
- 4: Geneseqp2001s:\*
- 5: Geneseqp2002s:\*
- 6: Geneseqp2003s:\*
- 7: Geneseqp2003bs:\*
- 8: Geneseqp2004s:\*
- 9: Geneseqp2005s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7901	100.0	1474	8	Adp18622 Photorhab
2	7901	100.0	1474	8	Adr21525 Photorhab
3	7285.5	92.2	1476	6	Abm70222 Photorhab
4	6065.5	76.8	1477	6	Abm70230 Photorhab
5	6042	76.5	1481	2	Aay33730 Photorhab
6	6031.5	76.3	1476	5	Abg32652 P. lumine
7	6025.5	76.3	1476	8	Adr21576 Photorhab
8	4627.5	58.6	1486	6	Abm70526 Photorhab
9	4595.5	58.2	1485	8	Adr21578 Photorhab
10	4581.5	58.0	1485	2	Aaw56546 Toxin Tca
11	4580.5	58.0	1485	2	Aaw17887 Photorhab
12	4261	53.9	1506	8	Adr21529 Xenorhabd
13	4261	53.9	1506	9	Aeb47812 Native Xp
14	3971	50.3	1428	4	Aay97695 Sepb prot
15	3971	50.3	1428	8	Adr21540 Serratia
16	3834	48.5	1493	8	Adr20369 Recombina
17	3834	48.5	1493	8	Adr21498 Xenorhabd
18	3106.5	39.3	1444	8	Adm61347 Paenibaci
19	3106.5	39.3	1444	8	Adr21520 Paenibaci
20	1744.5	22.1	662	3	Aay95685 Cosmid cH
21	1262	16.0	697	8	Adm61375 Paenibaci
22	839	10.6	591	2	Aar23006 Protein t
23	199	2.5	2060	5	Aae20967 Staphyloc
24	172.5	2.2	1404	7	Adc00960 Enterohae

25	170	2.2	593	6	AAO19846	Aao19846 P aerugin
26	169.5	2.1	14130	9	AEA07438	Aea07438 M. ulcera
27	162.5	2.1	16990	9	AEA07436	Aea07436 M. ulcera
28	161	2.0	1863	7	ADD46989	Add46989 Rat Prote
29	161	2.0	1863	9	ADx26456	Adx26456 Novel cel
30	160.5	2.0	1385	6	ABU40318	Abu40318 Protein e
31	160.5	2.0	1426	3	ABU15983	Abu15983 E. coli p
32	160.5	2.0	1426	6	ABU14693	Abu14693 Protein e
33	159.5	2.0	1329	4	ABG28402	Abg28402 Novel hum
34	159.5	2.0	1329	4	ABG25780	Abg25780 Novel hum
35	159.5	2.0	1329	7	ADC33172	Adc33172 Human nov
36	155.5	2.0	658	7	ABO76164	AbO76164 Pseudomon
37	155.5	2.0	1400	7	ADf01365	Adf01365 Enterohae
38	155.5	2.0	2315	7	ADP05149	Adp05149 Bacterial
39	154	1.9	4630	3	AAY77177	Aay77177 S. venezu
40	154	1.9	5215	8	ADL91930	AdL91930 Streptomy
41	153.5	1.9	1028	7	ADP06896	AdP06896 Bacterial
42	152.5	1.9	1669	8	ADU07870	AdU07870 Amino aci
43	152	1.9	4630	2	AAW19629	Aaw19629 Streptomy
44	151.5	1.9	1904	5	ABB57100	Abb57100 Mouse lsc
45	151.5	1.9	1904	9	ADx26385	Adx26385 Novel cel

#### ALIGNMENTS

#### RESULT 1

ID	ADP18622	standard; protein; 1474 AA.
XX	ADP18622	
AC	ADP18622;	
XX		
DT	12-AUG-2004	(first entry)
XX		
DE	Photorhabdus luminescens W-14	tcdB2 protein SEQ ID NO:10.
XX		
KW	tcd; Photorhabdus luminescens W-14;	transgenic plant; Toxin A;
KW	orally active insect toxin; insect toxin;	tcdB2.
XX		
OS	Photorhabdus luminescens.	
XX		
PN	WO2004044217-A2.	
XX		
FD	27-MAY-2004.	
XX		
PF	12-NOV-2003; 2003WO-IB005553.	
XX		
PR	12-NOV-2002; 2002US-0425672P.	
XX		
FA	(UYBA-) UNIV BATH.	
XX		
PI	Ffrench-Constant RH, Waterfield NR;	
XX		
DR	WPI; 2004-411735/38.	
DR	N-PSDB; ADP18621.	
XX		
PT	New isolated Photorhabdus luminescens nucleic acids, useful for	
PT	expressing orally active insect toxin or for generating transgenic plants	
PT	with enhanced resistance to insects.	
XX		
PS	Claim 1; SEQ ID NO 10; 118pp; English.	
XX		

The present invention describes DNA sequences from the tcd genomic region of Photorhabdus luminescens W-14. Also described: (1) a transgenic monocot or dicot cell having a genome comprising a nucleic acid sequence that encodes SEQ ID NO:4, 6, 8, 10, 12, 14, or 16; (2) a transgenic plant with a genome comprising a nucleic acid sequence that encodes SEQ ID NO:4, 6, 8, 10, 12, 14, or 16; (3) a seed of the transgenic plant; (4) a progeny of seed; (5) a method of producing Toxin A of Photorhabdus luminescens W-14 in a heterologous host; and (6) a method of producing an orally active insect toxin. The nucleotide sequences are useful for heterologous expression of orally active insect toxin. They can also be used for generating transgenic plants with enhanced resistance to

CC	insects. The present sequence represents Photorehabdus luminescens W-14																																																									
CC	tcdB2, which is used in the exemplification of the present invention.																																																									
XX																																																										
SQ	Sequence 1474 AA;																																																									
	Query Match	100.0%;	Score	7901;	DB	8;	Length	1474;																																																		
	Best Local Similarity	100.0%;	Pred. No.	0;																																																						
	Matches 1474;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;																																																	
Qy	1	MONSQPSITELSLPKGGGAI	TGGEAL	TP	TG	DG	MAALS	SL	PL	PISAGRGYAPAF	PL	NYN	60																																													
Db	1	MONSQPSITELSLPKGGGAI	TGGEAL	TP	TG	DG	MAALS	SL	PL	PISAGRGYAPAF	PL	NYN	60																																													
Qy	61	SGAGNSPGLGWCN	VMTIRRT	TH	FG	VP	HYDE	TD	TP	LG	PEGE	VL	VADQ	PR	DE	ST	LQ	GIN	120																																							
Db	61	SGAGNSPGLGWCN	VMTIRRT	TH	FG	VP	HYDE	TD	TP	LG	PEGE	VL	VADQ	PR	DE	ST	LQ	GIN	120																																							
Qy	121	LGATFTVTGYSRL	SHFSR	LEYW	QPKT	TG	TF	WLI	YSP	DG	QV	HL	LK	GSP	Q	A	R	I	SN	PS	Q	180																																				
Db	121	LGATFTVTGYSRL	SHFSR	LEYW	QPKT	TG	TF	WLI	YSP	DG	QV	HL	LK	GSP	Q	A	R	I	SN	PS	Q	180																																				
Qy	181	TTOTAQWLL	EA	SV	SR	GE	QI	YQ	YR	AE	DD	TG	CE	AD	EI	TH	HL	QA	Q	R	Y	L	H	I	V	Y	G	N	R	T	A	S	240																									
Db	181	TTOTAQWLL	EA	SV	SR	GE	QI	YQ	YR	AE	DD	TG	CE	AD	EI	TH	HL	QA	Q	R	Y	L	H	I	V	Y	G	N	R	T	A	S	240																									
Qy	241	ETLPGLDGS	APSO	ADW	L	F	Y	L	V	D	Y	G	E	R	S	N	L	K	T	P	P	A	F	S	T	T	G	S	W	L	C	R	O	D	R	F	S	R	Y	E	Y	G	P	E	I	R	300											
Db	241	ETLPGLDGS	APSO	ADW	L	F	Y	L	V	D	Y	G	E	R	S	N	L	K	T	P	P	A	F	S	T	T	G	S	W	L	C	R	O	D	R	F	S	R	Y	E	Y	G	P	E	I	R	300											
Qy	301	TRRLCRQV	LMYHL	QALD	S	K	I	T	E	H	G	P	T	L	V	S	R	L	I	N	Y	D	E	S	A	I	A	S	T	L	V	F	V	R	R	V	G	H	E	Q	D	G	N	V	360													
Db	301	TRRLCRQV	LMYHL	QALD	S	K	I	T	E	H	G	P	T	L	V	S	R	L	I	N	Y	D	E	S	A	I	A	S	T	L	V	F	V	R	R	V	G	H	E	Q	D	G	N	V	360													
Qy	361	VTLPPL	E	L	A	Y	Q	D	F	S	P	R	H	A	W	Q	M	D	V	L	A	N	F	A	N	T	O	R	W	Q	L	V	D	L	K	E	G	L	P	G	L	L	Y	Q	D	K	A	W	W	Y	R	420						
Db	361	VTLPPL	E	L	A	Y	Q	D	F	S	P	R	H	A	W	Q	M	D	V	L	A	N	F	A	N	T	O	R	W	Q	L	V	D	L	K	E	G	L	P	G	L	L	Y	Q	D	K	A	W	W	Y	R	420						
Qy	421	SAORL	G	I	G	S	D	A	V	T	W	E	R	Q	W	P	L	S	V	I	P	S	I	Q	S	N	A	S	L	V	D	I	N	G	D	Q	L	D	W	I	T	G	P	L	R	G	V	H	S	O	R	P	D	480				
Db	421	SAORL	G	I	G	S	D	A	V	T	W	E	R	Q	W	P	L	S	V	I	P	S	I	Q	S	N	A	S	L	V	D	I	N	G	D	Q	L	D	W	I	T	G	P	L	R	G	V	H	S	O	R	P	D	480				
Qy	481	GSWTR	F	T	P	L	N	A	L	P	V	E	T	H	P	R	A	Q	A	L	D	M	G	A	G	L	S	D	L	V	L	I	G	P	K	S	V	R	L	Y	A	N	T	R	D	G	F	A	K	G	D	V	Q	540				
Db	481	GSWTR	F	T	P	L	N	A	L	P	V	E	T	H	P	R	A	Q	A	L	D	M	G	A	G	L	S	D	L	V	L	I	G	P	K	S	V	R	L	Y	A	N	T	R	D	G	F	A	K	G	D	V	Q	540				
Qy	541	SGDIT	L	P	V	P	G	A	D	R	K	L	V	A	F	S	D	V	L	G	S	G	A	H	L	V	E	S	A	T	K	V	T	C	W	P	N	L	G	R	G	R	F	C	O	P	I	T	L	P	G	F	S	Q	600			
Db	541	SGDIT	L	P	V	P	G	A	D	R	K	L	V	A	F	S	D	V	L	G	S	G	A	H	L	V	E	S	A	T	K	V	T	C	W	P	N	L	G	R	G	R	F	C	O	P	I	T	L	P	G	F	S	Q	600			
Qy	601	PATF	E	N	P	A	Q	V	T	L	A	D	G	S	G	P	T	L	I	Y	H	T	N	R	L	D	I	F	L	N	K	S	G	N	G	F	A	P	E	P	V	T	L	R	P	E	G	L	R	P	D	H	T	C	660			
Db	601	PATF	E	N	P	A	Q	V	T	L	A	D	G	S	G	P	T	L	I	Y	H	T	N	R	L	D	I	F	L	N	K	S	G	N	G	F	A	P	E	P	V	T	L	R	P	E	G	L	R	P	D	H	T	C	660			
Qy	661	QLQ	M	A	D	V	Q	L	G	V	A	S	L	I	L	S	V	P	H	M	S	P	H	R	W	C	D	L	T	N	K	P	W	L	L	N	E	M	N	N	M	G	V	H	H	T	L	R	Y	S	S	S	Q	720				
Db	661	QLQ	M	A	D	V	Q	L	G	V	A	S	L	I	L	S	V	P	H	M	S	P	H	R	W	C	D	L	T	N	K	P	W	L	L	N	E	M	N	N	M	G	V	H	H	T	L	R	Y	S	S	S	Q	720				
Qy	721	WLDE	K	A	A	A	L	T	G	T	P	V	C	L	P	P	H	T	L	W	O	T	E	D	E	I	S	G	N	K	L	V	T	L	R	Y	A	R	G	A	D	G	R	E	R	F	R	G	780									
Db	721	WLDE	K	A	A	A	L	T	G	T	P	V	C	L	P	P	H	T	L	W	O	T	E	D	E	I	S	G	N	K	L	V	T	L	R	Y	A	R	G	A	D	G	R	E	R	F	R	G	780									
Qy	781	FGV	Y	E	Q	T	D	S	H	O	L	A	G	N	A	P	E	R	T	P	P	A	L	T	K	N	W	A	T	G	L	P	V	I	D	N	A	L	S	T	E	Y	W	R	D	D	O	A	F	A	G	F	S	P	R	F	T	840
Db	781	FGV	Y	E	Q	T	D	S	H	O	L	A	G	N	A	P	E	R	T	P	P	A	L	T	K	N	W	A	T	G	L	P	V	I	D	N	A	L	S	T	E	Y	W	R	D	D	O	A	F	A	G	F	S	P	R	F	T	840
Qy	841	TWQ	N	K	D	V	P	L	T	P	D	N	S	R	Y	F	N	R	A	L	K	Q	L	L	R	S	E	L	Y	G	L	D	S	T	N	K	H	V	P	T	V	T	E	F	R	S	Q	V	R	R	L	Q	900					
Db	841	TWQ	N	K	D	V	P	L	T	P	D	N	S	R	Y	F	N	R	A	L	K	Q	L	L	R	S	E	L	Y	G	L	D	S	T	N	K	H	V	P	T	V	T	E	F	R	S	Q	V	R	R	L	Q	900					
Qy	901	HTD	S	R	Y	P	V	L	M	S	S	V	E	S	R	N	Y	H	E	R	I	A	S	D	P	Q	S	O	N	I	T	L	S	S	D	R	F	Q	P	L	K	S	V	O	Y	P	R	Q	O	P	A	I	960					
Db	901	HTD	S	R	Y	P	V	L	M	S	S	V	E	S	R	N	Y	H	E	R	I	A	S	D	P	Q	S	O	N	I	T	L	S	S	D	R	F	Q	P	L	K	S	V	O	Y	P	R	Q	O	P	A	I	960					

Qy	961	NLY	P	D	T	L	P	D	K	L	A	N	S	Y	D	D	O	R	Q	R	L	T	Y	Q	O	S	S	W	H	L	T	N	T	V	R	V	L	G	L	P	D	S	T	R	S	D	I	F	T	Y	G	A	E	N	1020					
Db	961	NLY	P	D	T	L	P	D	K	L	A	N	S	Y	D	D	O	R	Q	R	L	T	Y	Q	O	S	S	W	H	L	T	N	T	V	R	V	L	G	L	P	D	S	T	R	S	D	I	F	T	Y	G	A	E	N	1020					
Qy	1021	VP	A	G	G	L	N	E	L	L	S	D	K	N	S	L	I	A	D	D	K	P	R	E	Y	L	G	O	O	K	T	A	Y	T	D	G	O	N	T	P	L	O	T	P	T	R	Q	A	L	I	A	F	T	T	V	F	1080			
Db	1021	VP	A	G	G	L	N	E	L	L	S	D	K	N	S	L	I	A	D	D	K	P	R	E	Y	L	G	O	O	K	T	A	Y	T	D	G	O	N	T	P	L	O	T	P	T	R	Q	A	L	I	A	F	T	T	V	F	1080			
Qy	1081	NQ	S	T	L	S	A	F	N	G	S	I	P	S	D	K	L	S	T	L	E	O	A	G	Y	O	O	T	N	Y	L	P	R	T	G	E	D	K	V	W	A	H	G	Y	T	D	Y	G	T	A	A	Q	F	W	R	P	1140			
Db	1081	NQ	S	T	L	S	A	F	N	G	S	I	P	S	D	K	L	S	T	L	E	O	A	G	Y	O	O	T	N	Y	L	P	R	T	G	E	D	K	V	W	A	H	G	Y	T	D	Y	G	T	A	A	Q	F	W	R	P	1140			
Qy	1141	QK	O	S	N	T	O	L	T	G	K	I	L	I	W	A	N	Y	C	V	V	Q	T	R	D	A	G	L	T	T	S	A	K	Y	D	W	R	F	L	T	P	V	Q	L	T	D	I	N	D	N	O	H	L	I	T	L	1200			
Db	1141	QK	O	S	N	T	O	L	T	G	K	I	L	I	W	A	N	Y	C	V	V	Q	T	R	D	A	G	L	T	T	S	A	K	Y	D	W	R	F	L	T	P	V	Q	L	T	D	I	N	D	N	O	H	L	I	T	L	1200			
Qy	1201	D	A	L	G	R	P	I	T	L	R	F	M	T	G	T	E	N	G	K	M	T	G	Y	S	S	P	E	K	A	S	F	S	P	S	D	V	N	A	A	I	E	L	K	K	L	P	V	A	Q	Q	Y	V	A	P	E	S	W	M	1260
Db	1201	D	A	L	G	R	P	I	T	L	R	F	M	T	G	T	E	N	G	K	M	T	G	Y	S	S	P	E	K	A	S	F	S	P	S	D	V	N	A	A	I	E	L	K	K	L	P	V	A	Q	Q	Y	V	A	P	E	S	W	M	1260
Qy	1261	P	V	L	S	O	K	T	F	N	R	L	A	B	O	D	W	K	L	N	A	R	I	I	T	E	D	G	R	I	C	T	L	A	Y	R	R	W	V	O	S	O	K	A	I	P	O	L	I	S	L	L	N	G	P	R	L	P	1320	
Db	1261	P	V	L	S	O	K	T	F	N	R	L	A	B	O	D	W	K	L	N	A	R	I	I	T	E	D	G	R	I	C	T	L	A	Y	R	R	W	V	O	S	O	K	A	I	P	O	L	I	S	L	L	N	G	P	R	L	P	1320	
Qy	1321	P	H	S	L	T	T	D	R	Y	D	H	D	P	Q	O	I	R	O	Q	V	F	S	D	G	F	G	R	L	O	A	A	R	H	E	A	G	H	A	R	O	R	N	E	D	G	S	L	I	I	N	V	O	H	T	1380				
Db	1321	P	H	S	L	T	T	D	R	Y	D	H	D	P	Q	O	I	R	O	Q	V	F	S	D	G	F	G	R	L	O	A	A	R	H	E	A	G	H	A	R	O	R	N	E	D	G	S	L	I	I	N	V	O	H	T	1380				
Qy	1381	E	N	R	W	A	T	G	R	E	Y	D	N	K	O	P	I	T	Y	O	P	F	L	N	D	W	R	Y	V	S	N	D	S	A	R	O	K	E	A	Y	A	D	T	H	V	Y	D	P	I	G	R	E	I	K	V	1440				
Db	1381	E	N	R	W	A	T	G	R	E	Y	D	N	K	O	P	I	T	Y	O	P	F	L	N	D	W	R	Y	V	S	N	D	S	A	R	O	K	E	A	Y	A	D	T	H	V	Y	D	P	I	G	R	E	I	K	V	1440				
Qy	1441	I	T	A	K	G	W	F	R	T	L	F	T	P	W	T	V	N	E	D	N	T	A	A	E	V	K	K	V	K	1474																													
Db	1441	I	T	A	K	G	W	F	R	T	L	F	T	P	W	T	V	N	E	D	N	T	A	A	E	V	K	K	V	K	1474																													
RESULT 2																																																												
ADR21525																																																												
ID ADR21525 standard; protein; 1474 AA.																																																												
XX	AC	ADR21525;																																																										
XX	XX																																																											
DT	04-NOV-2004	(first entry)																																																										
XX	XX																																																											
DE	DE	Photorhabdus TcdB2 toxin SEQ ID NO:45.																																																										
XX	XX																																																											
KW	KW	toxin; insect; insecticidal; transgenic; pest control.																																																										
XX	XX																																																											
OS	OS	Photorhabdus luminescens.																																																										
XX	XX																																																											
PN	PN	WO2004067727-A2.																																																										
PD	PD	12-AUG-2004.																																																										
XX	XX																																																											
PF	PF	07-JAN-2004; 2004WO-US000394.																																																										
XX	XX																																																											
PR	PR	21-JAN-2003; 2003US-0441723P.																																																										
XX	XX																																																											
PA	PA	(DOWC ) DOW AGROSCIENCES LLC.																																																										
XX	XX																																																											
PI	PI	Hey TD, Schleper AD, Bevan SA, Bintrim SB, Mitchell JC, Li ZS;																																																										
PI	PI	Ni W, Zhu B, Merlo DJ, Apel-Birchhold PC;																																																										
XX	XX																																																											
DR	DR	WPI; 2004-580999/56.																																																										
XX	XX	N-PSDB; ADR21524.																																																										
XX	XX																																																											
PT	PT	Controlling or inhibiting an insect, useful for pest control, comprises																																																										
PT	PT	contacting the insect with effective amounts of a Protein A, a Protein B,																																																										
XX	XX	and a Protein C.																																																										
XX	XX																																																											
PS	PS	Claim 1; SEQ ID NO 45; 368pp; English.																																																										

CC The invention relates to a novel method for controlling or inhibiting an  
 CC insect comprising contacting the insect with effective amounts of a  
 CC Protein A, a Protein B, and a Protein C. Each of the Proteins A, B, and C  
 CC is encoded by a naturally occurring gene or has an amino acid sequence  
 CC that differs from the product encoded by a naturally occurring gene only  
 CC by truncation or by conservative amino acid changes. Protein A is a 230-  
 CC 290 kDa toxin complex insect toxin that is derived from a first taxonomic  
 CC species, has stand alone insecticidal activity, and has an amino acid  
 CC sequence at least 40% identical to a sequence selected from XpA2w1,  
 CC XpA2w1, TcdA, TcdA2, TcdA4, and TcdA. Protein B is a 130-180 kDa toxin  
 CC complex potentiator having an amino acid sequence at least 40% identical  
 CC to a sequence selected from TcdB1, TcdB2, TcdC, XpC1w1, XpC1Bxb,  
 CC PptB1(orfs), or SepB. Protein C is a 90-120 kDa toxin complex potentiator  
 CC having an amino acid sequence at least 35% identical to a sequence  
 CC selected from TcdC1, TcdC2, TcdC3, TcdC4, TcdC5, XpC1w1, XpC1Bxb, PptC1  
 CC (orf 6 long), PptC1 (orf 6 short), and SepC. Also claimed is a transgenic  
 CC plant or plant cell that produces a Protein A, a Protein B, and a Protein  
 CC C. The method is useful for pest control. The present sequence represents  
 CC Photorhabdus luminescens TcdB2 toxin.

XX  
 SQ Sequence 1474 AA;

Query Match 100.0%; Score 7901; DB 8; Length 1474;  
 Beat Local Similarity 100.0%; Pred. No. 0;  
 Matches:1474; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MNSQDSFISITSLPKGGGATGMEALTPGPGMAALSPLPISAGRGYAPFTLNYN 60  
 Db 1 MNSQDSFISITSLPKGGGATGMEALTPGPGMAALSPLPISAGRGYAPFTLNYN 60  
 Qy 61 SGAGNSPGLGWCNVMTRIRTHFGVPHYDETDTFLGPEGEVLVADQPRDESTLOGIN 120  
 Db 61 SGAGNSPGLGWCNVMTRIRTHFGVPHYDETDTFLGPEGEVLVADQPRDESTLOGIN 120  
 Qy 121 LGATFTVTVGRSRLSHFSLRLEYQPKTTGKTDFWLIYSPDQVHLLGKSPQARISNPSQ 180  
 Db 121 LGATFTVTVGRSRLSHFSLRLEYQPKTTGKTDFWLIYSPDQVHLLGKSPQARISNPSQ 180  
 Qy 181 TTQTAQMLLEASVSRGEQIYYQRAEDDTGCEADEITHLQATQRYLHVITYGNRTAS 240  
 Db 181 TTQTAQMLLEASVSRGEQIYYQRAEDDTGCEADEITHLQATQRYLHVITYGNRTAS 240  
 Qy 241 ETLPLDGSAPSQADWLFYLPDYDGERSNLKTTPAFSTTGSWLCRODRFSRYEGFEIR 300  
 Db 241 ETLPLDGSAPSQADWLFYLPDYDGERSNLKTTPAFSTTGSWLCRODRFSRYEGFEIR 300  
 Qy 301 TRRLCROVLMYHLQALDSKITEHNGPTLVSRLLINVDSEIASTLVFVRVGHQDGNV 360  
 Db 301 TRRLCROVLMYHLQALDSKITEHNGPTLVSRLLINVDSEIASTLVFVRVGHQDGNV 360  
 Qy 361 VTLPPLLAYQDFSPRHHQPMQMDVLNFAIQRWLVDLKGELPGLLYQDKGAWMYR 420  
 Db 361 VTLPPLLAYQDFSPRHHQPMQMDVLNFAIQRWLVDLKGELPGLLYQDKGAWMYR 420  
 Qy 421 SAQRLGEGSDAVTWKMQPLSVIPSLQSNASLVNDINGDGLDWITGPGRLRGVHSORPD 480  
 Db 421 SAQRLGEGSDAVTWKMQPLSVIPSLQSNASLVNDINGDGLDWITGPGRLRGVHSORPD 480  
 Qy 481 GSWTRFTPLNALPVEYTHPRQAQDLGAGLSDLVLGPKSVRLYANTRDGFAGKQVQV 540  
 Db 481 GSWTRFTPLNALPVEYTHPRQAQDLGAGLSDLVLGPKSVRLYANTRDGFAGKQVQV 540  
 Qy 541 SGDITLPPVPGADPRKLNAFSDVLSGQAHLEVSATKVTWPNLGRGFRGFPITLPGFSQ 600  
 Db 541 SGDITLPPVPGADPRKLNAFSDVLSGQAHLEVSATKVTWPNLGRGFRGFPITLPGFSQ 600  
 Qy 601 PATEFNPAQVYLAIDGSGPTDLIVYHNRDIDIFLNKSGNGFASPVTLRPFEGRLRPHHC 660  
 Db 601 PATEFNPAQVYLAIDGSGPTDLIVYHNRDIDIFLNKSGNGFASPVTLRPFEGRLRPHHC 660  
 Qy 661 QLQWADVOGLGVASLIISLVPHMSPHHWRCDLTNNKPKLLNEMNNMNVGHHTLYRSSSQF 720  
 Db 661 QLQWADVOGLGVASLIISLVPHMSPHHWRCDLTNNKPKLLNEMNNMNVGHHTLYRSSSQF 720

Qy 721 WLDEKAAALTTGOTPVCYLPFPIHTLMQTEDEBSIGNKLVTTILRYARGAWDGREREPRG 780  
 Db 721 WLDEKAAALTTGOTPVCYLPFPIHTLMQTEDEBSIGNKLVTTILRYARGAWDGREREPRG 780  
 Qy 781 FGVEQTDHQLAGNAPERTPPALTKNWATGVLVDNALSTEVWRDDQAFAGSPRFT 840  
 Db 781 FGVEQTDHQLAGNAPERTPPALTKNWATGVLVDNALSTEVWRDDQAFAGSPRFT 840  
 Qy 841 TWQNDKDVLTPTDDNSRYFNRALKQALLRSELYGLDDSTNKHVPYTVTFRQVRLQ 900  
 Db 841 TWQNDKDVLTPTDDNSRYFNRALKQALLRSELYGLDDSTNKHVPYTVTFRQVRLQ 900  
 Qy 901 HTDSRYPLVSSVSVESRYHYERIASDIPQSCNITLSSDRFGQPLKOLSVQYPRQPAI 960  
 Db 901 HTDSRYPLVSSVSVESRYHYERIASDIPQSCNITLSSDRFGQPLKOLSVQYPRQPAI 960  
 Qy 961 NLXPDTLPDKLLANSYDDQORQLRLTYQSSWHHLTNTVVRVLGLPSTRSDIFTYGAEN 1020  
 Db 961 NLXPDTLPDKLLANSYDDQORQLRLTYQSSWHHLTNTVVRVLGLPSTRSDIFTYGAEN 1020  
 Qy 1021 VPAGLNLLELLSKNSLIADDKPREYLGQOKTAYTDGQNTTPTLQTRQALIAFTETVF 1080  
 Db 1021 VPAGLNLLELLSKNSLIADDKPREYLGQOKTAYTDGQNTTPTLQTRQALIAFTETVF 1080  
 Qy 1081 NOSTLSAFNGSIPSDKLSITLLEQAGYQOTNYLPPRTGEDKVVVAHGYTDYGTAAQFWRP 1140  
 Db 1081 NOSTLSAFNGSIPSDKLSITLLEQAGYQOTNYLPPRTGEDKVVVAHGYTDYGTAAQFWRP 1140  
 Qy 1141 OKQSNTOLTKKITLWDANYCVVVQTRDAAGLTTSKDYWRFLTPVOLTDINDNOHLITL 1200  
 Db 1141 OKQSNTOLTKKITLWDANYCVVVQTRDAAGLTTSKDYWRFLTPVOLTDINDNOHLITL 1200  
 Qy 1201 DALCRPTTLFAFWGTENGKMTGYSPEKASFPSPSDVNAALTELKPLPVAOCQVYAPSSWM 1260  
 Db 1201 DALCRPTTLFAFWGTENGKMTGYSPEKASFPSPSDVNAALTELKPLPVAOCQVYAPSSWM 1260  
 Qy 1261 PVLQKPTFNRLAODWOKLYNARIITEDGRICITLAYERVWQSKAIQPOLISLNNGPRLP 1320  
 Db 1261 PVLQKPTFNRLAODWOKLYNARIITEDGRICITLAYERVWQSKAIQPOLISLNNGPRLP 1320  
 Qy 1321 PHSITLTTDDYDHDPEQIRQOVVFSDFGRLLOAAARHEAGMARQORNEGSLINVOHT 1380  
 Db 1321 PHSITLTTDDYDHDPEQIRQOVVFSDFGRLLOAAARHEAGMARQORNEGSLINVOHT 1380  
 Qy 1381 ENRWAVTGRTEYDNKGQPIRTYQPYFLNDWRYVNSDSAROEKEAYADTHVYDPIGREIKV 1440  
 Db 1381 ENRWAVTGRTEYDNKGQPIRTYQPYFLNDWRYVNSDSAROEKEAYADTHVYDPIGREIKV 1440  
 Qy 1441 ITAKGWFRRLTFTPWFTVNEDENDTAAEVKKVKM 1474  
 Db 1441 ITAKGWFRRLTFTPWFTVNEDENDTAAEVKKVKM 1474

RESULT 3

ABM70222  
 ID ABM70222 standard; protein; 1476 AA.

XX AC ABM70222;

XX AC ABM70222;

XX AC ABM70222;

XX AC ABM70222;

XX AC ABM70222;

XX AC ABM70222;

XX AC ABM70222;

XX AC ABM70222;

XX AC ABM70222;

PD 28-NOV-2002.  
 XX  
 PF 07-FEB-2002; 2002WO-IB003040.  
 XX  
 PR 07-FEB-2001; 2001FR-00001659.  
 XX  
 PA (INSP ) INST PASTEUR.  
 PA (CNRS ) CNRS CENT NAT RECH SCI.  
 XX  
 XX  
 PI Duclaud E, Taourit S, Glaser P, Frangeul L, Kunst F, Danchin A;  
 PI Buchrieser C;  
 XX  
 XX WPI; 2003-148459/14.  
 XX  
 XX Genomic sequence of Photorhabdus luminescens and encoded polypeptides,  
 PT useful e.g. as therapeutic antimicrobials and agricultural pesticides.  
 XX  
 XX Claim 2; SEQ ID NO 3319; 1205pp; French.  
 XX  
 XX The invention relates to the isolation of genes and their encoded  
 CC proteins from Photorhabdus luminescens. The isolated sequences are  
 CC sources of probes and primers for detecting the genome of P. luminescens  
 CC and related species; to study polymorphisms; for gene analysis and for  
 CC detection/amplification of the genes. Antibodies (Ab) raised against the  
 CC polypeptides encoded by the genes are used for detection/identification  
 CC of P. luminescens, e.g. in foods. The genes, proteins, Ab and cells that  
 CC carry a gene-containing vector are used to select compounds that  
 CC modulate, regulate, induce or inhibit expression of the genes in plants,  
 CC animals or microorganisms other than P. luminescens and are able to alter  
 CC response or sensitivity to toxins and antibiotics produced by P.  
 CC luminescens. Cells transformed to express the genes are useful for  
 CC recombinant production of the proteins, particularly toxins and  
 CC antibacterials useful as insecticides, bactericides and fungicides. The  
 CC genes, proteins, vectors containing the genes and Ab are also useful  
 CC therapeutically (to treat microbial infection by bacteria or fungi that  
 CC are sensitive to P. luminescens-encoded toxins or antibiotics) and as  
 CC biopesticides. Other uses of the genes and the proteins are as virulence  
 CC factors and for identifying targets of human diseases for which P.  
 CC luminescens is a model (particularly plague and whooping cough). This  
 CC sequence represents one of the isolated P. luminescens proteins  
 XX  
 XX Sequence 1476 AA;  
 XX  
 Query Match 92.2%; Score 7285.5; DB 6; Length 1476;  
 Best Local Similarity 91.7%; Pred. No. 0;  
 Matches 1352; Conservative 55; Mismatches 67; Indels 1; Gaps 1;  
 QY 1 MNSQDPSITELSLPKGGGAIITGMEALTPGPDGMAALSPLPISAGRYAPAFITLNYN 60  
 DB 1 MNSQDPSITELSLPKGGGAIITGMEALTPGPDGMAALSPLPISAGRYAPAFITLNYN 60  
 QY 61 SGAGNSPFLGWCNVMVIRRRTHFGVPHYDETDFTFLGPEGEVLVADQPRDSTLQGIN 120  
 DB 61 SGAGNTFPLGWCNVMVIRRRTHFGVPHYNETDFTFLGPEGEVLVADQPRDSTLQGIN 120  
 QY 121 LGATFTVTGYSRLESFHSLEYWQPKTKGTDFMLIYSDGQVHLGKSPQARISNPQ 180  
 DB 121 LGATFTVTGYSRLESFHSLEYWQPKTKGTDFMLIYSDGQVHLGKSPQARISNPQ 180  
 QY 181 TTQTAQWLLBASVSRGEIYYQYRAEDDTCCEADEITHTLQAPQRYLHVYGNRTAS 240  
 DB 181 TTQTAQWLLBASVPHGEIYYQYRAEDNLDCTEITLHPQATQRYLHVYGNRTAS 240  
 QY 241 ETLPGLDGSAPSQADWLFLYLFVDFYGERSNLKPFPAPSTTGSMLCRODRFSRYEYGP 300  
 DB 241 ETLPGLNGRAPSQADWLFLYLFVDFYGERSNLKPFPAPSTTGSMLCRODRFSRYEYGP 300  
 QY 301 TRRLCROVLYMHLQALDSKITHEHGPTLVSRILNVDSEAIASLTIVFVRVGEQDGNV 360  
 DB 301 TRRLCROILMYHCLQALDNKIKHEHGPTLVSRILNVDSEAIASLTIVFVRVGEQDGT 360  
 QY 361 VTLPPLLELAYQDFSPRHHAWQPMQMDVLNFAIQRWQLVLDKGEGLPGLLYQDKGAWWYR 420  
 DB

DB 361 VTLPPLLELAYQDFSPQHNTDQPMQMDVLNFAIQRWQLVLDKGEGLPGLLYQDKGAWWYR 420  
 QY 421 SAQRLEIGSDAVTWKMQPLSVIPSLQSNASLVDINGOLDWVITPGLRQVHYSORPD 480  
 DB 421 SAQRLEIGSDVTVWEKIQLPSVPSQSNASLVDINGOLDWVITPGLRQVHYSORPD 480  
 QY 481 GSWTRFTPLNALPVEYTHPRAQLADLMGAGLSDLVLIGPKSVRLYANTRDGFAKGDVVQ 540  
 DB 481 GSWTRFTPLNALPVEYTHPRAQLADLMGAGLSDLVLIGPKSVRLYANTRDGFAKGDVVQ 540  
 QY 541 SGDTILPVPFADPRKLVAFSDVLGSGOAHLEVSATKVTCPNPLGRGRFQQTILPGFSQ 600  
 DB 541 SGDTILPVPFADPRKLVAFSDVLGSGOAHLEVSATKVTCPNPLGRGRFQQTILPGFSQ 600  
 QY 601 PATEFPAQVYLADLDGSGPTDLIYVHTNRDLIFLNKSGNGFAEPVTLRPEGLRPHDHC 660  
 DB 601 PEAIFPAQVYLADLDGSGPTDLIYVHTNRDLIFLNKSGNGFAEPVTLRPEGLRPHDHC 660  
 QY 661 QLMQADVQGLGVASLILSVPHMSPHHWRCDLTNMKPMLLNEMNNMGVHTTLAYRSSQF 720  
 DB 661 QLMQADVQGLGVASLILSVPHMTPHWRCDLTSAKPMLLSEMMNMGVHTTLAYRSSQF 720  
 QY 721 WLDEKAAALTGTQTPVCYLPFPPIHTLMQTEDEISGNKLVTTLRYARGAWDGRERFRG 780  
 DB 721 WLDEKAAALVAGTQTPVCYLPFPVHTLMQTEDEISGNKLVTTLRYARGAWDGRERFRG 780  
 QY 781 FGVEQTDHSHQAOGNAPERTPPALTKNMYATGLPIDNLTSTYWRGDNQAFAGSPRF 840  
 DB 781 FGVEQTDHSHQAOGNASERTPPAMTKNMYATGLPIDNLTSTYWRGDNQAFAGSPRF 840  
 QY 840 TTQDQNDKVPILTPEDDNSRYFNRLKGLLSRLSELGLDSTNKHVPYVTFERSQVRL 899  
 DB 841 TTQDQNDKVPILTPEDDNSRYFNRLKGLLSRLSELGLDSTNKHVPYVTFERSQVRL 900  
 QY 900 QHTDSRYPLVMSVSRNHYERIASDPQCSQNTLSSDRFQPKQLSVQVPRRQQA 959  
 DB 901 QQADNQYPLVMSVSRNHYERIASDPQCSQNTLSSDRFQPKQLSVQVPRRQQA 960  
 QY 960 INLYPDTLPDKLIANSYDDOQROLRLTYQSSWHHLTNNVTVRVLGPDSPRSIDFYGAE 1019  
 DB 961 ISYPDTLPDELLANSYDQROLRLTYQSSWHHLTNNVTVRVLGPDSPRSIDFYGAE 1020  
 QY 1020 NVPAGGLNLELLSDKNSLIADDPREYLGQOKTAYTDGQNTTPTQPTQALIAFTETTV 1079  
 DB 1021 NVPAGGLNLELLSDKNSLIADDPREYLGQOKTAYTDGQNTTPTQPTQALIAFTETTV 1080  
 QY 1080 FNOSTLSAFNGSTSPSKLSTLLEQAGYQQTNYLFPRTGEDKVVVAHHGYTDYGTAAQFWR 1139  
 DB 1081 FSQSTLSAFNGSTSPSKLSTLLEQAGYQQTNYLFPRTGEDKVVVAHHGYTDYGTAAQFWR 1140  
 QY 1140 PQKQSTOLTGKITLTDWANYCVVQTRDAAGLTTSKAYDMRELTPVQLTDINDNOHLIT 1199  
 DB 1141 PQKQSTOLTGKITLTDWANYCVVQTRDAAGLTTSKAYDMRELTPVQLTDINDNOHLIM 1200  
 QY 1200 LDALGRPITLRFWGTENGKMTGYSSPEKASFPSPDVNAIELKPLPVAQCQVYAPESW 1259  
 DB 1201 LDALGRPITLRFWGTENGKMTGYSSPEKASFPSPDVNSAIELKPLPVAQCQVYAPESW 1260  
 QY 1260 MPVLSOKTFNRLAEQDQWKLQYNARIITEGRICTLAYRRVWQSKAIPQLISLNNQPL 1319  
 DB 1261 MPVLSOKALNRLAEQDQWKLQYNARIITEGRICTLAYRRVWQSKATPQLTSLNNGPHL 1320  
 QY 1320 PPHSLTLTDRYDHDPEQRIQOVVSDGFRLLLOAAARHEAGMARQNEGSLINVOH 1379  
 DB 1321 PPHSLTLTDRYDHDPEQRIQOVVSDGFRLLLOAAARHEAGMARQNEGSLINVOH 1380  
 QY 1380 TENRNVATGRTYDNDKQPIRTYQPYFLNDWRVYVNSDSARQEKAEADTHVYDPVGREIK 1439  
 DB 1381 TENRNVATGRTYDNDKQPIRTYQPYFLNDWRVYVNSDSARQEKAEADTHVYDPVGREIK 1440  
 QY 1440 VITAKGFRRLTLPFTVNEDENDTAAAEVVKVYM 1474  
 DB 1441 VITAKGFRRLTLPFTVNEDENDTAAAEVNEVM 1475



## RESULT 4

ABM70230  
ID ABM70230 standard; protein; 1477 AA.

XX AC ABM70230;

XX XX 20-NOV-2003 (first entry)

XX DE Photorhabdus luminescens protein sequence #3327.

XX KW Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;  
XX KW detection; food; gene expression; plant; animal; microorganism; toxin;  
XX KW antibiotic; biopesticide; virulence factor; disease model; plague;  
XX KW whooping cough.

XX OS Photorhabdus luminescens.

XX PN W0200294867-A2.

XX XX 28-NOV-2002.

XX PF 07-FEB-2002; 2002WO-IB003040.

XX XX 07-FEB-2001; 2001FR-00001659.

XX PA (INSP ) INST. PASTEUR.

XX PA (CNRS ) CNRS CENT NAT RECH SCI.

XX PI Duchaud E, Taourit S, Glaser P, Frangeul L, Kunst F, Danchin A;  
XX PI Buchrieser C;

XX DR WPI; 2003-148459/14.

XX PT Genomic sequence of Photorhabdus luminescens and encoded polypeptides,  
XX PT useful e.g. as therapeutic antimicrobials and agricultural pesticides.

XX PS Claim 2; SEQ ID NO 3327; 1205pp; French.

XX CC The invention relates to the isolation of genes and their encoded  
XX CC proteins from Photorhabdus luminescens. The isolated sequences are  
XX CC sources of probes and primers for detecting the genome of P. luminescens  
XX CC and related species; to study polymorphisms; for gene analysis and for  
XX CC detection/amplification of the genes. Antibodies (Ab) raised against the  
XX CC polypeptides encoded by the genes are used for detection/identification  
XX CC of P. luminescens, e.g. in foods. The genes, proteins, Ab and cells that  
XX CC carry a gene-containing vector are used to select compounds that  
XX CC modulate, regulate, induce or inhibit expression of the genes in plants,  
XX CC animals or microorganisms other than P. luminescens and are able to alter  
XX CC response or sensitivity to toxins and antibiotics produced by P.  
XX CC luminescens. Cells transformed to express the genes are useful for  
XX CC recombinant production of the proteins, particularly toxins and  
XX CC antibacterials useful as insecticides, bactericides and fungicides. The  
XX CC genes, proteins, vectors containing the genes and Ab are also useful  
XX CC therapeutically (to treat microbial infection by bacteria or fungi that  
XX CC are sensitive to P. luminescens-encoded toxins or antibiotics) and as  
XX CC biopesticides. Other uses of the genes and the proteins are as virulence  
XX CC factors and for identifying targets of human diseases for which P.  
XX CC luminescens is a model (particularly plague and whooping cough). This  
XX CC sequence represents one of the isolated P. luminescens proteins

XX SQ Sequence 1477 AA;

Query Match 76.8%; Score 6065.5; DB 6; Length 1477;  
Best Local Similarity 76.3%; Pred. No. 0;  
Matches 1126; Conservative 120; Mismatches 221; Indels 9; Gaps 3;

Qy 1 MNSQDFSIITSLPKGGGAIITGMEALTPTGPDGMAALSPLPISAGRGYAPFTLNYN 60

Dd 1 MNSQTFVSELSLPLKGGGAIITGMEALTPTGPDGMAALSPLPISAGRGYSPSLTLYS 60

Qy 61 SGAGNSPFLGWCNVMVIRRTTHFGVPHYDETDTFLGPEGEVLVA-----DQPRDES 114

Db 61 SGAGNSPFLGWCNVMVIRRTTHFGVPHYDETDTFLGPEGEVLVAIALNENGQADISES 120  
Qy 115 TLQGINLGATFTVTGYSRLESFHSRLEYWQPKTKTKTDFWLIYSPDQVHLLGKSPQAR 174  
Db 121 SLQGINLGEFTVTGYSRLESFHSRLEYWQPKTKTKTDFWLIYSPDQVHLLGKSPQAR 180  
Qy 175 ISNPSQTTQAWMLLEASVSRSRGEIYYQRAEDDTCCEADEITHHQAQRIYLVHYY 234  
Db 181 ISNPLNVSTQAWMLLEASVSRSRGEIYYQRAEDDTCCEADEITHHQAQRIYLVHYY 240  
Qy 235 GNTASFTLGLDGSAPSQADWLFYLDYGERSNLKTTPAPSTTSGWLCRCRDRFRSYE 294  
Db 241 GNTASFTLGLDGSAPSQADWLFYLDYGERSNLKTTPAPSTTSGWLCRCRDRFRSYE 300  
Qy 295 YGFEIRTRRLCRQVLMYHHLQALDSKITENHNGPLTVSRLLILNYDESIASTLVFVRVGH 354  
Db 301 YGFEIRTRRLCRQVLMYHHLQALDSKITENHNGPLTVSRLLILNYDESIASTLVFVRVGH 360  
Qy 355 EQDGNVVTLPPLLEYAQDFSPRHHAHWPMQDVLANFNAIQRWQLVDLKGEGLPGLLYQDK 414  
Db 361 EDNNTVISLPLLEYAQDFSPRHHAHWPMQDVLANFNAIQRWQLVDLKGEGLPGLLYQDK 420  
Qy 415 GAWYRSRQRLGEIGSDAVTWKMOPLSVIPSLQSNASLVLDINGDGLDWVITGPGLRGY 474  
Db 421 NGWYRSRQRLGEIGSDAVTWKMOPLSVIPSLQSNASLVLDINGDGLDWVITGPGLRGY 480  
Qy 475 HSQRPDGSWTRFTPLNALPVEYTHPRAQLADLMAGLSDLVLIGPKSVRLYANTRDGFAP 534  
Db 481 HSQRPDGSWTRFTPLNALPVEYTHPRAQLADLMAGLSDLVLIGPKSVRLYANTRDGFAP 540  
Qy 535 GKDVVQSGDITLPPVGADPRKLVAFSVILSGQAHLYEVSATKVTCPNLRGRFGQPIT 594  
Db 541 GRDVVQSGDITLPPVGADPRKLVAFSVILSGQAHLYEVSATKVTCPNLRGRFGQPIT 600  
Qy 595 LKPGSQPATEFNAQVYLDLDCGSGPTDLIVHTNRDLDFLNKSGNGFAEPVTLRFPEGL 654  
Db 601 LKPGSQPATEFNAQVYLDLDCGSGPTDLIVHTNRDLDFLNKSGNGFAEPVTLRFPEGL 660  
Qy 655 RFDTTCQLQVADVQGLGAVSLLSVPHWPHWRCDLTNAKPNLLSENNMNGAHTLHY 714  
Db 661 RFDTTCQLQVADVQGLGAVSLLSVPHWPHWRCDLTNAKPNLLSENNMNGAHTLHY 720  
Qy 715 RSSSQFQLDEKAAALTTGQTPVCYLPPIHTLMQTEDEISGNKLVTLRYAAGWDGR 774  
Db 721 RSSSQFQLDEKAAALTTGQTPVCYLPPIHTLMQTEDEISGNKLVTLRYAAGWDGR 780  
Qy 775 EREPRGFGYVEQTDHQAQNAPEPTPALTKNMYATGLPIDNALSTEYWR-DDQAFA 833  
Db 781 EREPRGFGYVEQTDHQAQNAPEPTPALTKNMYATGLPIDNALSTEYWR-DDQAFA 840  
Qy 834 GFSRFTTWDQNDKQVPLTPDDNSRYWFRNALKGQLLSRSELYGLDDSTNKHVPYTVTFER 893  
Db 841 GFSRFTTWDQNDKQVPLTPDDNSRYWFRNALKGQLLSRSELYGLDDSTNKHVPYTVTFER 900  
Qy 894 SQVRLQHTDSRVPVLMSSVVSRYHYRIASDPQCSQNTLSSDRFGQPKQLSQVQYP 953  
Db 901 PQVRQLQDGTATSPVLWASVVRNRYHYRIASDPQCSQNTLSSDRFGQPKQLSQVQYP 960  
Qy 954 RROQPAINLYPDTLPDKLLANSYDDQORQLRTYQSSWHHLTNNTVRLGLPDRSDI 1013  
Db 961 RROQPAINLYPDTLPDKLLANSYDDQORQLRTYQSSWHHLTNNTVRLGLPDRSDI 1020  
Qy 1014 FTYGAENVVAGLNLLELLSDKNLSIADDKPREYLGQCKTAYTDGQNTTLPQTPTROALIA 1073  
Db 1021 FTYGAENVVAGLNLLELLSDKNLSIADDKPREYLGQCKTAYTDGQNTTLPQTPTROALIA 1080  
Qy 1074 FTETTFVNSQTSLSAFNGSIPSDKLSSTTLEAGYQQTNYLFPRTGEDKVVWAHGYDTYGT 1133  
Db 1081 FTETTFVNSQTSLSAFNGSIPSDKLSSTTLEAGYQQTNYLFPRTGEDKVVWAHGYDTYGT 1140  
Qy 1134 AAQWRPKQSNTOITQKILILWPDANCVVVQVTFDAAGLTTSKDYDRFLTPVQLTIND 1193



Db 1141 EAQFWRPVQAQRNTLLTGKTTLQMDTHYCVITQTQDAAGLTVALYDWRFLTPVQLTDIND 1200

Qy 1194 NOHLITLDALGRPTLRFWCTGKMTGYSSPEKASFPSPDYNAATELKKPLPVAQCQV 1253

Db 1201 NVHLITLDALGRPTLRFWCTGKMTGYSSPEKASFPSPDYNAATELKKPLPVAQCQV 1260

Qy 1254 YAPESWMPVLSQKTFNRLAEPQWQKLYNARIITEDGRICTLAYRRWVQSKAIPQLISLL 1313

Db 1261 YAPDSWMPVLSQKTFNRLAEPQWQKLYNARIITEDGRICTLAYRRWVQSKAIPQLISLL 1320

Qy 1314 NNGPRPLPHSLTUTTDYRDHPBQOIRQVVFSDGFGRLLOAAARHAGMARQNEGSL 1373

Db 1321 TNSIGLPPHNLTLTPDRYDRDSGQIHOQVAFSDGFGRLLOAAARHAGMARQNEGSL 1380

Qy 1374 IINVQHTENWAVTGRTEYDNKQPIRTYQPYFLNDWRYVNSDARQEKAYADTHYDP 1433

Db 1381 VTKMEDTKTWAVTGRTEYDNKQPIRTYQPYFLNDWRYVNSDARQEKAYADTHYDP 1438

Qy 1434 IGRKIVITAKGWRFRRLTFTPMFVNEDENDTAAEV 1469

Db 1439 IGRIRVITAKGMLRQSQYFPWFTVSEENDTAAEV 1474

RESULT 5

AAV33730

ID AAV33730 standard; protein; 1481 AA.

XX

AC AAV33730;

XX

XX 09-NOV-1999 (first entry)

XX

XX Photorhabdus luminescens 1481 amino acid insecticidal toxin.

XX

XX Symbiotic bacterium; nematode; insect; larva; toxin; insecticide.

XX

XX Photorhabdus luminescens.

XX

XX WO9942589-A2.

XX

XX 26-AUG-1999.

XX

XX 18-FEB-1999; 99WO-EP001015.

XX

XX 20-FEB-1998; 98US-00027080.

XX

XX 20-JAN-1999; 99US-0116439P.

XX

XX (NOVS ) NOVARTIS AG.

XX

XX (NOVS ) NOVARTIS-ERFINDUNGEN VERW GES MBH.

XX

XX Kramer VC, Morgan MK, Anderson AR, Hart HP, Warren GW, Dunn MW;

PI Chen JS;

XX

XX WPI; 1999-527479/44.

DR N-PSDB; AA206831.

XX

XX New nucleic acid from Photorhabdus luminescens encoding insecticidal

PT toxins, used for making resistant transgenic plants.

XX

XX Claim 26; Page 140-145; 148pp; English.

XX

XX This sequence represents a 1481 amino acid insecticidal toxin from

CC Photorhabdus luminescens. It is one of three insecticidal toxins

CC (AAV33728-Y33730) encoded by open reading frames (ORFs) in a 39kb

CC fragment of P. luminescens DNA (AAZ06831). This sequence is encoded by

CC orf2. P. luminescens is a member of the Enterobacteriaceae family and is

CC a symbiotic bacterium of nematodes of the genus Heterorhabditis. The

CC nematodes colonise insect larvae, kill them, and their offspring feed on

CC the dead larvae. However, the insecticidal agents are produced by P.

CC luminescens rather than the nematodes. The toxins have activity against

CC Lepidopteran insects such as Cabbage looper (*Trichoplusia ni*), European

CC Corn Borer (*Ostrinia nubilalis*) and Fall Armyworm (*Spodoptera frugiperda*)

CC and also against Coleopteran insects (e.g., Colorado Potato Beetle,

CC *Leptinotarsa decimlineata*). In addition the toxins are active against

CC strains resistant to known insecticides. The DNA sequence can be used to

CC generate transgenic plants of various species that are resistant to

CC economically important insect pests and also for recombinant production

CC of the toxins for use as insecticides

XX

SQ Sequence 1481 AA;

Query Match 76.5%; Score 6042; DB 2; Length 1481;

Best Local Similarity 76.1%; Pred. No. 0;

Matches 1127; Conservative 113; Mismatches 226; Indels 14; Gaps 5;

Qy 1 MONSQFSPITELSLPGGGAITGCGEALTPTGPDGMAALSPLPIPSAGRGYAPFTLYN 60

Db 1 MONSQFSPITELSLPGGGAITGCGEALTPTGPDGMAALSPLPIPSAGRGYAPFTLYN 60

Qy 61 SGAGNSPFGLGWDCNVTIRRTTHFGVPHYDETDFTLGPBEVGVVA-----DQPRDES 114

Db 61 SGAGNSPFGLGWDCNVTIRRTTHFGVPHYDETDFTLGPBEVGVVA-----DQPRDES 120

Qy 115 TLOGINLGATFTVTGYSRLESHPFSLRYEYQPKTKGTDFMLIYSPDQVHLHGKSPQAR 174

Db 121 SLOGINLGATFTVTGYSRLESHPFSLRYEYQPKTKGTDFMLIYSPDQVHLHGKSPQAR 180

Qy 175 ISNPSQTTTQAOMLLEASVSSRGEIYYQYRAEDDTGCEADEITHHLQATAQRYLHIVY 234

Db 181 ISNPLVNVQTAQWMLLEASVSSHGEIYYQYRAEDDTGCEADEITHHLQATAQRYLHIVY 240

Qy 235 GNRTASSETLPGLDGSAPSQADWLPLYVDFYGERSNMLKTPPAFTSTGSMWLCRODRFSRYE 294

Db 241 GNLTASEVFTPLNGDDPLKSGWLFCLVDFYGERKNSLSEMPFPKATSNMLCRKDRFSRYE 300

Qy 295 YGFEIRTRRLCROVLMYHHLQALDSDKITEHNGPTLVSRLLNYDESAIATSLVPRVRGH 354

Db 301 YGFALTRRLCROVLMYHHLQALDSDKITEHNGPTLVSRLLNYDESAIATSLVPRVRGH 360

Qy 355 EQDG-NVVTLPPLLELAYQDFSPRRHAWQPMVDYLANFNATQRLQVLDLKGELPGLLYQD 413

Db 361 EQDGTTAVALPPELAYQDFPEPEQKALMRPMDVLANFNATQRLQVLDLKGELPGLLYQD 420

Qy 414 KGAWYRSQRLGEIGSDAVTWBKQPLSVIPSLQSNASLVDINGDQLDQVITGPGLRG 473

Db 421 KNGWYRSQRLGEIGSDAVTWBKQPLSVIPSLQSNASLVDINGDQLDQVITGPGLRG 480

Qy 474 YHSORPDGSMTRFTPLNALPVEYTHPRAQLADLWAGAGLSDVLIGPKSVELYANTRDGEA 533

Db 481 YHSQHPDGSMTFTPLHALPIETHPRAQLADLWAGAGLSDVLIGPKSVELYANTRDGEA 540

Qy 534 KGKDVVQSGDITLTPFGADPRKLVAFSDVLGSGQAHLEVSATKVTCPNLRGRFGQPI 593

Db 541 EGRDVVQSGDITLTPFGADPRKLVAFSDVLGSGQAHLEVSATKVTCPNLRGRFGQPI 600

Qy 594 TLPFGSQPATEFNPAAQVYLADLQSGPTDLIYVHTNRDLIFLNKSGNGPAEPVTLRPPBG 653

Db 601 TLPFGSQSAAANFNPDRVHLADLQSGPADLIYVHADHLDIFSNEGNGPAQPTLRFPDG 660

Qy 654 LRFDHTCOLQADVOGLGVASLTLISVPHMSPHWRCDLTNMKPILLNEMNNMGVHHTLR 713

Db 661 LRFDHTCOLQADVOGLGVASLTLISVPHMSPHWRCDLTNMKPILLNEMNNMGVHHTLR 720

Qy 714 YRSSSFMLDEKAAALTTGTPVCYLPFPFIHTLWQTEDEISGNKLVTTLRYARGAWDG 773

Db 721 YRSSVFMLDEKAAALTTGTPVCYLPFPFVHTLWQTEDEISGNKLVTTLRYARGAWDG 780

Qy 774 REREPFGVGVTEQDSHQLAQGNAPERTPPALTKNMYATGLPVIDNALSTEYWR-DDQAF 832

Db 781 REREPFGVGVTEQDSHQLAQGNAPERTSPALTKNMYATGIPVDNLTSLAGYVRGDTQAF 840

Qy 833 AGSPRFTTWQDNKDVPLTPEDDNSRYWFRALKGOLLRSLEYGLDDSTNKHVPTVTFE 892

Db 841 TGTPIHFTLWKGKDVPLTPEDDNLNLRALKGQPLRSELYGLDGSAAQKIPYTVTES 900

Qy 893 RSQVRRLQHTDSRYPYPLWGSVVSERNHYERIASDPQCQNITLSSDRFOQPLKQISVQY 952

Db 901 RPQVRLQDNNTTSLPVLWASVSRSHYHRIISDPQCNDITLSSDLFGQPLKQSVQY 960  
 QY 953 PRQOPAINLYPDITLDPKLLANSYDDQORQLRLTYQSSWHLLTNTVTVLGLPDSRSD 1012  
 Db 961 PRRKPTTNPYPDTLPDITLFPASSYDQOQLRLTYQSSWHLLTNTVTVLGLPDSRSD 1020  
 QY 1013 IFTYGAENVPAGGNNLELLSKNSLIADDPREYLGQOQKATY----TDQNTTLPLOTPT 1068  
 Db 1021 AFTYDAKHVPDGLNLEALCAENSLIADDPREYLNQOQRTYTDGKTGKPTPLKPTPT 1080  
 QY 1069 QALIAFTETTVFNQNTLSAFNGSPSKLSTLLEQAGYQOQNTYLPFRGTGDKVVAHGY 1128  
 Db 1081 QALIAFTETAVLTSLISAFDGGITPDDELPGLLTQAGYQOQRTYLPFLSGENQVWARKY 1140  
 QY 1129 TDYGTAAQFWRQSQNTQLTKTLIWDANYCVVQTRDAAGLTTSKAYDWRFLTPVOL 1188  
 Db 1141 TDYGTGVEQFWRPVAQRNTQLTKTLKWDTHYCVITQTDAGLTVSANYDWRFLTPVOL 1200  
 QY 1189 TDINDNOHLITLDALGRPTITLRFWGTENGKWTGYSSPEKASFSPSDVNAALIELKPLPV 1248  
 Db 1201 TDINDNVHIIITLDALGRPTVQRFNGIENGVAIGSSPEAKPFTPPVDVNAALITGLPLV 1260  
 QY 1249 AQCVYAPESWMPVLSQKTFNRLAEQDMOKLYNARIITEDGRICITLAYRRWVQSKAIPQ 1308  
 Db 1261 AQCLVYAPDSWMPVLSQKTFNRLAEQDMOKLYNARIITEDGRICITLAYRRWVQSKAIPQ 1320  
 QY 1309 LISLNNQPRLLPHSLTTLTDYDHPDQOQRTYQOQVFSDFGRLLQAAARHEAGMARON 1368  
 Db 1321 LVKLTNLSIGLPHNLMATDRYDRDSQOQRTYQOQVFSDFGRLLQAAARHEAGMARON 1380  
 QY 1369 EDGSLIINVQHTENRWAVGTEDYDNKQPIRTYQOQVFSDFGRLLQAAARHEAGMARON 1428  
 Db 1381 QDGLSVTKQEDTKRWAITGTEDYDNKQPIRTYQOQVFSDFGRLLQAAARHEAGMARON 1448  
 QY 1429 HYPDIPGRIKIVITAKGFRITLTPFTVNEDENDTAAE 1468  
 Db 1439 HIYDIPGRIKIVITAKGFRITLTPFTVNEDENDTAAE 1478  
 RESULT 6  
 ID ABG32652 standard; protein; 1476 AA.  
 AC ABG32652;  
 XX 30-DEC-2002 (first entry)  
 DT 30-DEC-2002 (first entry)  
 XX P. luminescens (W-14) TcdB protein.  
 DE Gene; ds; toxin A; toxin B; TcdA; protoxin; TcdB; TccC2; transgenic;  
 KW monocot cell; dicot cell; oral toxin; insect; pest; TcdA.  
 XX Photothhabdus luminescens.  
 XX US2002078478-A1.  
 XX 20-JUN-2002.  
 XX 26-MAR-2001; 2001US-00817514.  
 XX 24-MAR-2000; 2000US-0191806P.  
 XX (FFRE/) FFRENCH-CONSTANT R H.  
 PA (BOWE/) BOWEN D.  
 PA (ROCH/) ROCHELEAU T A.  
 PA (WATE/) WATERFIELD N R.  
 XX Ffrench-Constant RH, Bowen D, Rocheleau TA, Waterfield NR;  
 PI WPI; 2002-655379/70.  
 XX N-PSDB; ABS52584.  
 XX Novel nucleic acid sequences which encode genes, tcdB and tccC2 from

PT Photothhabdus luminescens W-14, useful in heterologous expression of  
 XX orally active insect toxins.  
 XX Claim 1; Page 20-23; 40pp; English.  
 PS The invention discloses an isolated nucleic acid that encodes TcdB or  
 CC TccC2 from Photothhabdus luminescens W-14. Also disclosed is a transgenic  
 CC monocot or dicot cell and a transgenic plant (including the seeds) both  
 CC with genomes comprising tcdB and tccC2 nucleic acids. The nucleic acids  
 CC are useful for producing toxin A or B of P. luminescens W-14 in a  
 CC heterologous host and for encoding TcdB or TccC2 for producing an orally  
 CC active insect toxin in a host, where the host also expresses TcdA or TccB  
 CC from P. luminescens W-14. Heterologous expression of Toxin A does not  
 CC afford the level of oral toxicity to insects as that of the native toxin,  
 CC but the coexpression increases this toxicity. The transgenic plants  
 CC expressing effective amounts of the toxins protect themselves from insect  
 CC pests. When the insects feeds on the transgenic plant it also ingests the  
 CC toxins and this deters the insect from further biting into the plant and  
 CC may even harm or kill the insect. The sequence presented is the P.  
 CC luminescens (W-14) TcdB protein  
 XX  
 SQ Sequence 1476 AA;  
 Query Match 76.3%; Score 6031.5; DB 5; Length 1476;  
 Best Local Similarity 75.9%; Pred. No. 0;  
 Matches 1120; Conservative 122; Mismatches 224; Indels 9; Gaps 3;  
 QY 1 MNSQDSFISITELSLPKGGGAIATGMEALTPGPDGMAALSPLPISAGRGYAPFTLNNY 60  
 DB 1 MNSQDSFISITELSLPKGGGAIATGMEALTPGPDGMAALSPLPISAGRGYAPFTLNNY 60  
 QY 61 SCAGNSPFLGWDGCVNMTIRRTTHFGVPHYDETDFLGEPEGLVVA-----DQPRDES 114  
 DB 61 SGTGNSPFLGWDGCVNMTIRRTTHFGVPHYDETDFLGEPEGLVVA-----DQPRDES 120  
 QY 115 TLOGINLGATFTVGYRSRLESFHSRLEYWOPKTTGKDFWLIYSPDQVHLLGKSPQAR 174  
 DB 121 SLOGINLGATFTVGYRSRLESFHSRLEYWOPKTTGKDFWLIYSPDQVHLLGKSPQAR 180  
 QY 175 ISNPSQTTQTAQWLLEASVSRSRGEIYYQVRAEDDTGCEADEITHHQAQRYLHIVY 234  
 DB 181 ISNPLNVNQAQWLLEASVSRSRGEIYYQVRAEDDTGCEADEITHHQAQRYLHIVY 240  
 QY 235 GNRATSETLPGLDGSAPSOADWLFYVDFYGERSNLKPAPSTTSGWLCRODRFSRYE 294  
 DB 241 GNLATSDVFPFTLNGDDPLKSGMFCFLVDFYGERKNSLSEMPFLFKATGKWLCKDRFSRYE 300  
 QY 295 YGPEIRTRLCROVLMYHHLQALDSKITEHNGPTLVSRLLIYNDESAITLAVFRRVGH 354  
 DB 301 YGPEIRTRLCROVLMYHHLQALDSKITEHNGPTLVSRLLIYNDESAITLAVFRRVGH 360  
 QY 355 EQDGNVVTLPPELAYQDFSPRRHHAHQMDVLANFNAIQRWQLVLDLKGEGPLGLLYQDK 414  
 DB 361 EDNNTVTALPPELAYQDFSPRRHHAHQMDVLANFNAIQRWQLVLDLKGEGPLGLLYQDK 420  
 QY 415 GAWYRSQRLGEIGSDAVTWKMQPLSVIPSLQSNASLVNDINGDGLDWTGPGRLGY 474  
 DB 421 NGWYRSQRLGEIGSDAVTWKMQPLSVIPSLQSNASLVNDINGDGLDWTGPGRLGY 480  
 QY 475 HSQPDGSGWTRPTPLNALPVEYTHPRAQLADLMGAGLSDLVLIQPKSVRLYANTRDGEAK 534  
 DB 481 HSQPDGSGWTRPTPLNALPVEYTHPRAQLADLMGAGLSDLVLIQPKSVRLYANTRDGEAK 540  
 QY 535 GKDVVQSGDITLPVPGADPRKLVAFSDVLSGQAHVVEVSATKVTCPNLRGRFGQPI 594  
 DB 541 GRDVVQSGDITLPVPGADPRKLVAFSDVLSGQAHVVEVSATKVTCPNLRGRFGQPI 600  
 QY 595 LPGFSQPAFENPAQVYLADLDSGPTDLIYVHTNRDLIFLNKSGNGFAEPVTLRFPGL 654  
 DB 601 LPGFSQPAFENPAQVYLADLDSGPTDLIYVHTNRDLIFLNKSGNGFAEPVTLRFPGL 660  
 QY 655 RPDITCQLQADVQGLGVASLIISVPHMSPHHRCDLTNMKPWLNNNNNGVHTLRY 714

Db 661 RPDTCQLQVADYQGLGVVSLILSVPHMAPHHWRCDLTNAKPWLLSETNNMGNATHLY 720  
 QY 715 RSSQFWLDEKAAALTTGQTPVCYLPPPIHTLWQTEDEISGNKLVITLRYARGAWDGR 774  
 Db 721 RSSVQFWLDEKAAALATGQTPVCYLPPVHTLWQTEDEISGNKLVITLRYARGAWDGR 780  
 QY 775 EREFRFGVGYEQDTSHQACQAPERTPPALTGNWATGLPVDNALSTYWR-DQQAFA 833  
 Db 781 EREFRFGVGYEQDTSHQACQAPERTPPALTGSWATGLPAVDNALSGYWGDKQAFA 840  
 QY 834 GFSFRFTWQNDKVDPLTPEDDNRSYWFRNALGQLLRSELYGLDSTNKHVPTTYTEFR 893  
 Db 841 GFTFRFTLWKGKDVPLTPEDDNLWLNALGQLLRSELYGLDGAQOQIPVTVTESR 900  
 QY 894 SQVRLQHTSRPVLWSSVVERNHYERIASDPQCSNITLSSDRFGQPLKLSVQYYP 953  
 Db 901 PQVRLQDQATVSPVLWASVVERSYHYERIIISDPQCNQDITLSSDLFGQPLKLSVQYYP 960  
 QY 954 RROQPALNLYPDTLPDKLLANSYDDQORQLRLTVQSSWHHLTNWTVRVLGLPDRSRDI 1013  
 Db 961 RRNKPTNYPDTLPDTLPASSYDDQQLRLTCRQSSWHHLTGNELRVLGLPDRSRDA 1020  
 QY 1014 FTYGAENVAGGLNELLSKSLIADDPREYLGQOKTAYTDQONTPLQTPTRQALIA 1073  
 Db 1021 FTYDAQVVDGLNELTCAENSLIADDPREYLNQRTFTYDGNQOTPLKTPTRQALIA 1080  
 QY 1074 FTETTVNQSTLAFNGISPSDKSLSTLEAGYQOQNYLFPRTGEDKWWAHGYTDYGT 1133  
 Db 1081 FTETAVLTESLLSAFDGGITPDELPGILTQAGYQOQEPYLPFRTEGNKVMYARQGYTDYGT 1140  
 QY 1134 AAFQWRPQKQNTQKTILINDANYCVVVQTRDAAGLTTSYAKYDWRELTTPVQLTDIND 1193  
 Db 1141 EAQFWRVPAQRNLSLTGNTLKWDTYCVITQTDAAGLTVSANYDWRFTPTQLTDIND 1200  
 QY 1194 NQHLITLDALGRPITLFWGTENGKMTGYSSPEKASFSPSDVNAALTELKPLPVAQCV 1253  
 Db 1201 NVHLITLDALGRPVQTRFMGIESGVATGYSSSEKPFPPNDIDTALNLTGPLPVAQCV 1260  
 QY 1254 YAPESNMPVL-SQTFNRLAQDWQKLYNARIITEDGRICHTLAYRWVQSOKAIPOLISLL 1313  
 Db 1261 YAPDSNMPLSQETFNLTQOEGETLDRSRIITEDMWRICALTRRWLQSQKISTPLVKLL 1320  
 QY 1314 NNGPRLPPHSLITLTDYRDHDEQOIRQOVVPSDGFGLLQAAARHAGMARORNEPQSL 1373  
 Db 1321 TNSIGLPPHNLITLTDYRDSEQOIRQOVAFSDGFGLLQASVREHAGMARQNEPQSL 1380  
 QY 1374 IINVQHTENRWAVTGRTEYDNKGQPIRTYQPYFLNDWRVYVNSDARQEKAYADTHYDP 1433  
 Db 1381 VTKVENTKRWAVTGRTEYDNKGQPIRTYQPYFLNDWRVYVSDSAR--KEAYADTHYDP 1438  
 QY 1434 IGREIKVITAKGWFRRTLTPFWFTVNEDENDTAAE 1468  
 Db 1439 IGREIRVITAKGLRQSQYFPFWFTVSEDENDTAAE 1473

## RESULT 7

ADR21576  
 ID ADR21576 standard; protein; 1476 AA.

AC ADR21576;  
 XX  
 DT 04-NOV-2004 (first entry)  
 XX  
 DE Photorhabdus TcdB1 toxin.  
 XX  
 KW toxin; insect; insecticidal; transgenic; pest control.  
 XX  
 OS Photorhabdus luminescens.  
 XX  
 PN W02004067727-A2.  
 XX  
 PD 12-AUG-2004.  
 XX

PF 07-JAN-2004; 2004WO-US0000394.  
 XX  
 PR 21-JAN-2003; 2003US-0441723P.  
 XX  
 PA (DOWC) DOW AGROSCIENCES LLC.  
 XX  
 PI Hey TD, Schleper AD, Bevan SA, Bintrim SB, Mitchell JC, Li ZS;  
 PI Ni W, Zhu B, Merlo DJ, Apel-Birkhold PC;  
 XX  
 DR WPI; 2004-580999/56.  
 DR N-PSDB; ADR21502.  
 XX  
 PT Controlling or inhibiting an insect, useful for pest control, comprises  
 PT contacting the insect with effective amounts of a Protein A, a Protein B,  
 PT and a Protein C.  
 XX  
 PS Claim 1; SEQ ID NO 22; 368pp; English.  
 XX  
 CC The invention relates to a novel method for controlling or inhibiting an  
 CC insect comprising contacting the insect with effective amounts of a  
 CC Protein A, a Protein B, and a Protein C. Each of the Proteins A, B, and C  
 CC is encoded by a naturally occurring gene or has an amino acid sequence  
 CC that differs from the product encoded by a naturally occurring gene only  
 CC by truncation or by conservative amino acid changes. Protein A is a 230-  
 CC 230 kDa toxin complex insect toxin that is derived from a first taxonomic  
 CC species, has stand alone insecticidal activity, and has an amino acid  
 CC sequence at least 40% identical to a sequence selected from XptAlwi,  
 CC XptA2wi, TcdA, TcdA2, TcdA4, and TcdB. Protein B is a 130-180 kDa toxin  
 CC complex potentiator having an amino acid sequence at least 40% identical  
 CC to a sequence selected from TcdB1, TcdB2, TcdC, XptC1wi, XptB1xb,  
 CC PotB1(orf5), or SepB. Protein C is a 90-120 kDa toxin complex potentiator  
 CC having an amino acid sequence at least 35% identical to a sequence  
 CC selected from TcdC1, TcdC2, TcdC3, TcdC4, TcdC5, XptB1wi, XptC1xb, PptC1  
 CC (orf 6 long), PptC1 (orf 6 short), and SepC. Also claimed is a transgenic  
 CC plant or plant cell that produces a Protein A, a Protein B, and a Protein  
 CC C. The method is useful for pest control. The present sequence represents  
 CC Photorhabdus luminescens TcdB1 toxin.  
 XX  
 SQ Sequence 1476 AA;

Query Match 76.3%; Score 6025.5; DB 8; Length 1476;  
 Best Local Similarity 75.9%; Pred. No. 0;  
 Matches 1119; Conservative 122; Mismatches 225; Indels 9; Gaps 3;  
 QY 1 MQNSQPSITELSLPKGGGAIITGCEALTPGDMALSLPLPISAGRYAPAFITLNYN 60  
 DB 1 MQNSQPSITELSLPKGGGAIITGCEALTPGDMALSLPLPISAGRYAPAFITLNYN 60  
 QY 61 SGAGNSPFGLGWDCNVMTIRRRTHFGVPHYDETDITFLGPEGEVLVVA-----DQPRDES 114  
 DB 61 SGTGNSPFGLGWDCGVNAIRRTSTGVPNYDETDITFLGPEGEVLVVALEAGQADIRSES 120  
 QY 115 TLOGINLGATFTVTGYRSRLESFHSRLEYWQPKTKTDFWLIYSPDQGVHLLGKSPQAR 174  
 DB 121 SLOGINLGATFTVTGYRSRLESFHSRLEYWQPKTKTDFWLIYSPDQGVHLLGKSPQAR 180  
 QY 175 ISNPSQTTQTAQWILLEASVSSRGEQIYYQYRAEDDTGCEADEITHLQAQRYLHVY 234  
 DB 181 ISNPLVYNQTAQWILLEASVSSRGEQIYYQYRAEDDEAGCETDELAHAFSATVQRYLQVHY 240  
 QY 235 GNRITASETLPLGDSAPSQADWLFYLVDFYGERSNLKTPTTAFSTTGSWLCRCQDRFSRYE 294  
 DB 241 GNLITASDVFTFLNGDDPLKSGWMLFCLVDFYGERKNSLSEMPFKATGNWLCRCQDRFSRYE 300  
 QY 295 YGFEIRTRRLCRQVLMYHHLQALDSKITHEHPTLSRLILNYDESIASTLVFVRVGH 354  
 DB 301 YGFEIRTRRLCRQVLMYHHLQALDSKITHEHPTLSRLILNYDESIASTLVFVRVGH 360  
 QY 355 EQDGNVVTLPLPPLAYQDFSPRHAWQPMQDVLANFNAIQRWQLVDLKGGLPLGLLYQDK 414  
 DB 361 EDNNTVTLPLPPLAYQDFSPRHAWQPMQDVLANFNAIQRWQLVDLKGGLPLGLLYQDR 420  
 QY 415 GAWWYRSQRIGETGSDAVTWKMQPLSVIPSLQSNASLVLDINGDGLDVTITGFLRGY 474









DR	WPI; 1998-179427/16.	QY	651	PEGLRPHDHTCOLQADVQGLGVASLILSVPHMSPHHRCOLTNMKPWLNNMNNNNVHH	710
XX	N-PSDB; AAV29924.	Db	660	PEGVQDNTCOLQVADIQSLGASLILTPHAPHHRCOLSLTKPALLNVMNNNGAHH	719
PT	Isolated toxins from Photorhabdus luminescens strains - useful for	QY	711	TLYRSSQFVLDEKAAALTTQOTPVCIYLPFPPIHTLMQOTETDEISGNKLVTLLRARGA	770
PT	control of insect pests.	Db	720	TLYRSSAQFVLDEKQLTKAGSPACYPFPFPHMLLWYTEIQDEISGNRLTSEVNSHG	779
XX	Claim 34; Page 196-200; 321pp; English.	QY	771	WDGRRPFGVGVVEQTDHSHQLAQNAPERTPALTKNWVATGLPVIDNALSTEYWR-DD	829
CC	The present sequence represents a protein named TcaC of the bacterium	Db	780	WDGRRPFGVGVVEQTDHSHQLAQNAPERTPALTKNWVATGLPVIDNALSTEYWR-DD	839
CC	Photorhabdus luminescens (W-14). This is a symbiotic bacterium of the	QY	830	QAPAGSPRFTTQ--DNKDVPLTPBDDNSRVNFRNALKQILRSILYGLDDSTNKHVPY	887
CC	nematodes of the Heterorhabditis genus. The bacterium has at least 4	Db	840	QAYSGETRYTWHDNTQDQAFTP-NETQRNWLTRALKQILLRTLYGLDGTDKQTPY	898
CC	distinct genomic regions, tca, tcb, tcc, and tcd. Peptide products are	QY	888	TVTFPSQVRRLQHTDSRYPVLWSSVVEGRNHYHYRIASDPQCSQNTISSDRFGAPLKQ	947
CC	produced from these regions that are associated with insecticidal	Db	899	TVSESRQVRSIPVNETELSAWVTAIENRSVHYERIITDPQFSQIKLQHDIFGQSLSQ	958
CC	activity. The native toxins are secreted proteins. The proteins are toxic	QY	948	LSVQYPRRQCPALNLYPDTLPDKLLANSYDDQORQLRTVQSSMHLTNNTVTVLGLPD	1007
CC	to insects upon exposure and especially when ingested. The nucleic acid	Db	959	VDIAPRRRPAVNPYPPTLPETLPDSYDDQOQLRLVRQKNMHLTDGENWRLGLPN	1018
CC	sequence can be used to produce transgenic plants, baculoviruses or	QY	1008	STRSDIFTYCAENVNVPAGLNLELLSDKNSLIADDKPREVLGQOQKATYATDQNTTPTQPT	1067
CC	microbial hosts for toxin production. They can be used to control insects	Db	1019	AQRDVTYVDRSKIPTGEGISLEILLKDDGLLADKAAVILGQOQTFYTAGAEVLEKPT	1078
CC	pests from the Lepidoptera, Coleoptera, Hymenoptera, Diptera,	QY	1068	RQALIAFTETTVNQSTLSAFNGSIPIKSTLLEQAGYQOQNTYLPFRGTGDKVVAHNG	1127
CC	Dictyoptera, Acarina or Homoptera orders, especially the Southern or	Db	1079	LQALVAFQETAMMDTSLQAYEGVIEBEQELNLTALTOAGYQOQVRLNTRSESPVMAARQ	1138
CC	Western corn rootworm, Colorado potato beetle, mealworm, boll weevil,	QY	1128	YTDYGTAAQFWRQKQSNLTQKTLIWDANYCVVVTQDAAGLTTSKYDWRFLTPVQ	1187
CC	turf grub, beetle armyworm, black cutworm, cabbage looper, codling moth,	Db	1139	YTDYGTAAQFWRQKQSNLTQKTLIWDANYCVVVTQDAAGLTTSKYDWRFLTPVQ	1198
CC	corn earworm, European corn borer or tobacco hornworm or budworm	QY	1188	LTDINDNQHLITLDALGRPITLRFMGTEGKMTGYSPKASPSPPSDVNAALIELKKPLP	1247
XX	Sequence 1485 AA;	Db	1199	LTDINDNQHLITLDALGRPITLRFMGTEGKMTGYSPKASPSPPSDVNAALIELKKPLP	1255
Query Match	58.0%; Score 4581.5; DB 2; Length 1485;	QY	1248	VAQCVYAPESMMPVLSQKTFNRL---ABQDWKLYNARIITEDGRICTLAYRWVQSOK	1304
Best Local Similarity	57.6%; Pred. No. 0;	Db	1256	VAQCVYAPESMMPVLSQKTFNRL---ABQDWKLYNARIITEDGRICTLAYRWVQSOK	1315
Matches	858; Conservative 224; Mismatches 378; Indels 29; Gaps 11;	QY	1305	AIQQLISLNNGRPLPHSLTLTDRYDHDPEQOIRQOQVPSFGRLGLQAAARHAGMA	1364
1	MONSDFSITELSLPKGGGATGMGEALTPGPDGMAALSPLPISAGRYAPAFITLYN 60	Db	1316	LTIQLISLNNGRPLPHSLTLTDRYDHDPEQOIRQOQVPSFGRLGLQAAARHAGMA	1375
1	MODSPEVSIITLSPKGGGAINMGALNAAGPDGMAALSPLPISAGRYAPAFITLYN 60	QY	1365	RQNEDEGLIINQ-----HTENRWAVTGRTEYDNKGOPRTYQYFLNDWRYVNSDA	1418
61	SGAGNSPFLGWCNVMTIRRTHFVPHYDETDFTFLGPEGEVLVA-----DQPRDE 113	Db	1376	WQKEDGLVVDANGVLVSAPTDRNAVSGTEYDDKQPVRTYQYFLNDWRYVNSDA	1435
61	NSAGNGPFGIGWCGVMSISRRTHQGIPOYQNDOTFLSPQGEVNMIALNDQGPDIRQDV 120	QY	1419	RQKEAYADTHVVDPIGREIKVITAKGMFRFTLTFTWFTVNEDENDTAA 1467	
114	STIQINGLNGATFTVGRSLESHPFLSLEYWQPKT--TGKTDWFLIYSPDQVHLGKSP 171	Db	1436	RD--DLFADTHLYDPLGREYKIVITAKKLYREKLYTPWFVISEDENDTAS 1482	
121	KTLQGVTLPISYVTRVQARQLDFSKIEVWQASQGEGR-FWLISDPDGHILGKTA 179	QY	AAW17887		
172	QARISNPSTQTAQWLLASVSRGEQIYYQYRAEDDTGCEADEITHLQATAQRYLHI 231	Db	AAW17887;		
180	QACLANFQNDQIAQWLLSLEETVTPAGEHVSQYRAEDEAHCDNEKTAHPNVTAQRYLVQ 239	QY	17-OCT-2003 (revised)		
232	VYGNRTASTLPGDGSASQADWLFYVFDYDERSNNLKTTPAPST-TGSHLCRQDRF 290	Db	29-JAN-1998 (first entry)		
240	VNYGNIKPQASLFVLDNAPPAPEWFLVFDHGERDTSLSHTVPTWDAGTAQMSVPRDIF 299	QY	Photorhabdus luminescens insect toxin protein TcaC.		
291	SRVEYGEIETRRLLRCOVLMYHLQALDSKITEHNGPTLVSRLLIYDESAIASTLVFVR 350	Db	Insecticide; insect; toxin; pest control; biological control;		
300	SRVEYGEVETRRLLRCOVLMYHLQALDSKITEHNGPTLVSRLLIYDESAIASTLVFVR 359	QY	Photorhabdus luminescens; TcaC; Southern corn rootworm;		
351	RVGHEQDGNVVTLPPLSLAYQDFSPRHHAHQMPDMVLNFAINAIQWOLVDLKGSLPGLL 410	Db	Colorado potato beetle; Western corn rootworm; meal worm; boll weevil;		
360	QLSHESDGRPVTPQPPLELAWQRPDLEKIPVQWQFDALDNFNSQOQYQLVLRGSLPGML 419	QY	turf grub; Coleoptera; beet armyworm; black cutworm; cabbage looper;		
411	YQDKGAWYSQAQRLGIGSDAVTWKQPLSVTPSLQSNASLVYDINGDQOLDWITGPG 470	Db	codling moth; corn earworm; European corn borer; tobacco hornworm;		
420	YQDKGAWYSQAQRLGIGSDAVTWKQPLSVTPSLQSNASLVYDINGDQOLDWITGPG 479	QY			
471	LRGHSORPDGSTRFTPLNALPVEYTHPRAQLADLMGAGLSDLVLIGPKSVRLYANTRD 530	Db			
480	IRGHSOQPDGKTHFTFIPALPVEYTHPRAQLADLMGAGLSDLVLIGPKSVRLYANTRD 539	QY			
531	GFAGKDVQSGDITLVPVGPADPKVAFSDVLGSGQAHLVEVSATKVTCPNIGRGRPG 590	Db			
540	GWRKGEDVPQSTGITLPTVTGTDAKLVAFSDMLGSGQOHLVEIKGNRVTCPNIGRGRPG 599	QY			
591	QPTLPGFQSPATFPNPAQVYLADLDSGPTDILYVHTNRDLDFLNKSGNGFAPVTLRF 650	Db			
600	QPTLPGFQSPATFPNPAQVYLADLDSGPTDILYVHTNRDLDFLNKSGNGFAPVTLRF 659	QY			

RESULT 11

AAW17887

ID AAW17887 standard; protein; 1485 AA.

XX

AC

XX

DT

XX

DE

XX

KW

KW

KW

KW



KW tobacco budworm; Lepidoptera; Hymenoptera; Diptera, Dictyoptera; Acarina; Homoptera.

XX OS Photorhabdus luminescens; strain W-14 (ATCC 55397).

XX PN WO9717432-A1.

XX PD 15-MAY-1997.

XX PF 06-NOV-1996; 96WO-05018003.

XX PR 06-NOV-1995; 95US-0007255P.

XX PR 28-FEB-1996; 96US-00608423.

XX PR 28-AUG-1996; 96US-00705484.

XX (WISC) WISCONSIN ALUMNI RES FOUND.

XX PA Ensign JC, Bowen DJ, Petell J, Fatig R, Schoonover S;

XX PI Ffrench-Constant RH, Rocheleau TA, Blackburn MB, Hey TD, Merlo DJ;

XX PI Orr GL, Roberts JL, Strickland JA, Guo L, Cliche T;

XX DR WPI; 1997-281022/25.

XX DR N-PSDB; AAT68840.

XX PT Photorhabdus sp. insecticidal protein toxins and DNA encoding them - can

XX PT be genetically engineered into insect larvae food and plants for insect

XX XX control.

XX PS Claim 34; Page 159-164; 276pp; English.

XX CC This polypeptide comprises a specifically claimed insecticidal toxin

XX CC protein, TcaA, of Photorhabdus luminescens, a 184 kDa component of a

XX CC toxin protein complex. TcaB can be expressed in host cells using a gene

XX CC (see AAT68837) isolated from a genomic library. Claimed toxin proteins of

XX CC P. luminescens (see AAW17871, AAW17884-89, AAW17899-900, AAW18301-06) can

XX CC be applied to, or genetically engineered into, insect larvae food and

XX CC plants for insect control. The Photorhabdus toxins are particularly

XX CC effective against southern corn rootworm, Colorado potato beetle, Western

XX CC corn rootworm, meal worm, boll weevil and turf grub (Coleoptera), beet

XX CC armyworm, black cutworm, cabbage looper, codling moth, corn earworm,

XX CC European corn borer, tobacco hornworm and tobacco budworm (Lepidoptera),

XX CC and are also active against insects of the orders Hymenoptera, Diptera

XX CC Dictyoptera, Acarina and Homoptera. (All claimed). (Updated on 17-OCT-

XX CC 2003 to standardize OS field)

XX SQ Sequence 1485 AA;

Query Match 58.0%; Score 4580.5; DB 2; Length 1485;

Best Local Similarity 57.6%; Fred. No. 0;

Matches 858; Conservative 224; Mismatches 378; Indels 29; Gaps 11;

Qy 1 MONGDPSITELSLPKGGGALTGMEALTPPTGPGMAALSPLPISAGRGVAPAFITLNYN 60

Db 1 MDSPEVSIITLSPKGGGALNGEALNAAGPDGMAASLPLPISAGRGVAPAFITLNYN 60

Qy 61 SGAGNSPFLGMDCNVMTIRRTFHGPHYDETDTFLGPEGEVLVVA-----DQPRDE 113

Db 61 NSAGNGPFGICWQGVMSISRTTGIGIPQYGNDDTFLSPQGEVNMALNDQGPDIRQDV 120

Qy 114 STLQGINLGATFTVTGYSRLSHESRLEYKQPKT--TGKTDFTMLIYSPDQGVHLGKSP 171

Db 121 KTLQGVTLPISTYTRYQARQILDFSKIEYVQPASGQEGRA-FWLISLTPDGLHLGNTA 179

Qy 172 QARTSNPQVTTQATWLEASVSSRGEQIYYQYRAEDDTGCEADBITHLQATQRYLHI 231

Db 180 QACLANPQNDQILQALWLEETVTPAGEHVSQYRAEDBAHCDNEKTAHFNLTQRYLVQ 239

Qy 232 VYGNRTASETLPLGDSAPSQADWLFYLVFDYDYSRNLKTPPAFST-TGSMCLQRDRF 290

Db 240 VNYGNKPKQASLFLVDNAPPAPPEWLPHLVFDHGERDTSHTVPTWDAGTAQMSVRPDI 299

Qy 291 SRYEGFPIRTRRLCRQVLMTHHQLQALDSKITEHNGPTLVSRLLINVDSEAIASITLVFR 350

Db 300 SRYEGFPIRTRRLCRQVLMTHHQLQALDSKITEHNGPTLVSRLLINVDSEAIASITLVFR 359

Qy 351 RVGHEQDGNVVTLPPLPPLAYQDFSPRHHAHQMDVLANFNAIQRWQLVLDKSGELPCLL 410

Db 360 QLSHESDGRPVTPPPLPPLAYQDFSPRHHAHQMDVLANFNAIQRWQLVLDKSGELPCLL 419

Qy 411 YQDKGAWYRSQRLGEGISDAVTEWKNQPLSVIPSLQSNASLVLDINGDQGLDQVITGPG 470

Db 420 YQDRGAWYKAPQRBQEDGSDNAVYTKIAPLPLPPLQDNASLVLDINGDQGLDQVITGPG 479

Qy 471 LRGYHSORPDGSWTRFTPLNALPVEYTHPRAQLADLMGAGLSLDLVLPKSPVRLYANTRD 530

Db 480 IRGYHSQOPDGKTHFTFIPNALPVEYTHPRAQLADLMGAGLSLDLVLPKSPVRLYANTRD 539

Qy 531 GFAKGDVWQSGDITLPPVPGADPRKLVAFSDVLGSGQAHLVEVSATKVTCPNLRGRFRG 590

Db 540 GWRKGEDVPQSTGITLPTGTGTDARKLVAFSDMLGSGQOHLVEIKGNRVTCPNLRGRFRG 599

Qy 591 QPITLPGFSQAPATFPAQVYLADLDSGPTDLIYVHTNRLDIPLNKSGNGFABPVTIRP 650

Db 600 QPITLPGFSQAPATFPAQVYLADLDSGPTDLIYVHTNRLDIPLNKSGNGFABPVTIRP 659

Qy 651 PEGLRPDHTCOLQWADVOGLGVASLILSVPHMSPHWRCDLTNNKPVLLNEMNNMVGHH 710

Db 660 PEGVQFDNTCOLQWADVOGLGVASLILSVPHMSPHWRCDLTNNKPVLLNEMNNMVGHH 719

Qy 711 TLRYRSSQFWLDEKAAALTTGQTPVCYLPFPPIHTLMQTEDEISGNKLVITLRYARGA 770

Db 720 TLRYSSAQFWLDEKAAALTTGQTPVCYLPFPPIHTLMQTEDEISGNKLVITLRYARGA 779

Qy 771 WDGRERFRGFGVVEQDTHQLAGGNAPERTPPALTKNWTATGLPVIDNALSTELTYR-DD 829

Db 780 WDGRERFRGFGVVEQDTHQLAGGNAPERTPPALTKNWTATGLPVIDNALSTELTYR-DD 839

Qy 830 QAPAGSPRTTQW--DNKDVPLTPEDDNRVYFNFRALKGOLLSELYGLDDSTNKVVPY 887

Db 840 QAPAGSPRTTQW--DNKDVPLTPEDDNRVYFNFRALKGOLLSELYGLDDSTNKVVPY 898

Qy 888 TVTEFRSQVRLQHTDSRYVPLWSSVVESSRYRYERIASDPQCSQNTILSSDREGQLPKQ 947

Db 899 TVSESRVQVRSIPVKNKTELSSAWTAIENRSYHYERITDPQRSQSIKLQHDIFGQSLSQS 958

Qy 948 LSVQVPRQOPAINLYPDTLPDKLLANSYDDQORQLRLTYQQSSWHHLTNTVTVLGLPD 1007

Db 959 VDIAMPREKPAVNPYPPTLFTLFDSSYDDQQLLRLVRQKNSWHHLTDCENWRLGLPN 1018

Qy 1008 STSDIFTYGAENVPAAGLNLLELLSDKNLSIADDPREYLGQOKTATYDQNTTPTLQPT 1067

Db 1019 AQRRDVTYDRSKIPTTGISLEILLKDDGLLADKAAVILGQQQTFYTAGQAEVTLKPT 1078

Qy 1068 RQALIAFTTTFVFNQSTLSAFNGSIPSDKLSLTLLEQAGYQQTNYLPFRGTGDKVVAHGG 1127

Db 1079 LQALVAFQETAMDDNSLQAYEGVIEBQELNTELTQAGYQVAVLFWTRSESFWAAARQ 1138

Qy 1128 YTDYGTAAQFWRPQKQNTOLTGKILTIWDANYCVVQTRDAAGLTTSKADYMRFLTPVQ 1187

Db 1139 YTDYGTAAQFWRPQKQNTOLTGKILTIWDANYCVVQTRDAAGLTTSKADYMRFLTPVQ 1198

Qy 1188 LTIINDNHLITLDALGRPITLFWGTENGKMTGYSPEKASFPSPDVAALTELKPLP 1247

Db 1199 LTIINDNHLITLDALGRPITLFWGTENGKMTGYSPEKASFPSPDVAALTELKPLP 1255

Qy 1248 VAOCQVYAPESWMPVLSQKTFNRL--AEQDWOKLYNARIITEDGRITCLAYRWVSQK 1304

Db 1256 VAOCQVYAPESWMPVLSQKTFNRL--AEQDWOKLYNARIITEDGRITCLAYRWVSQK 1315

Qy 1305 AIPQLISLNNKGRPLPHSLTLTTRDYDHPDEQIIRQOVVFSFGFGLLQAAARHEAGMA 1364

Db 1316 LTIQLISLNASIPRLPHVLGITTDYDSDPQQHQQTQVFSFGFGLLQAAARHEAGMA 1375

Qy 1365 RORNESSLINQV-----HTENRMAVTGRTEYDNKGPITRYQPYFLNDWRVYSDSA 1418

Db 1376 WORKEDGGLVVDANGVLVLSAPTDTTRMAVSGRTEYDDKGPVRYQPYFLNDWRVYSDSA 1435

**Qy** 1419 RQEAYADTHVYDPIGREIKVIITAKGWFRRTLTPWTFVNEDENDTAA 1467  
| :  
**Db** 1436 RD--DLFADTHLYDPLGRYKVIITAKLYLREKLTPWTFIVSENDNTAS 1482

**RESULT 12**

ADR21529  
ID ADR21529 standard; protein; 1506 AA.

AC ADR21529;

DT 04-NOV-2004 (first entry)

Xenorhabdus toxin XptB1xb SEQ ID NO:49.

KW toxin; insect; insecticidal; transgenic; pest control.

OS Xenorhabdus bovienii.

PN WO2004067727-A2.

PD 12-AUG-2004.

PF 07-JAN-2004; 2004WO-US0000394.

PR 21-JAN-2003; 2003US-0441723P.

PA (DOWC ) DOW AGROSCIENCES LLC.

PI Hey TD, Schleper AD, Bevan SA, Bintrim SB, Mitchell JC, Li ZS;

[illegible]

DR N-PSDB; ADR21528.

Controlling or in

2.4 Connecting the antibody with catalase amounts of a protein A, PT and a Protein C.

PS Claim 1; SEQ ID NO 49; 368pp; English.

The invention relates to a novel method for controlling or inhibiting an insect comprising contacting the insect with effective amounts of a protein A, a protein B, and a protein C. Each of the proteins A, B, and C is encoded by a naturally occurring gene or has an amino acid sequence that differs from the product encoded by a naturally occurring gene only by truncation or by conservative amino acid changes. Protein A is a 230-290 kDa toxin complex insect toxin that is derived from a first taxonomic species, has stand alone insecticidal activity, and has an amino acid sequence at least 40% identical to a sequence selected from XptAlwI, XptA2wI, TcdA, TcdA2, TcdA4, and TcdB. Protein B is a 130-180 kDa toxin complex potentiator having an amino acid sequence at least 40% identical to a sequence selected from TcdB1, TcdB2, TcAc, XptC1wI, XptB1b, XptC1b1(or15), or SepB. Protein C is a 90-120 kDa toxin complex potentiator having an amino acid sequence at least 35% identical to a sequence selected from TccC1, TccC2, TccC3, TccC4, TccC5, XptB1wI, XptC1b, PptC1 (orf 6 long), PptC1 (orf 6 short), and SepC. Also claimed is a transgenic plant or plant cell that produces a protein A, a protein B, and a protein C. The method is useful for pest control. The present sequence represents *Xenorhabdus bovienii* XptB1b.

Sequence 1506 AA;

Query Match 53.9%; Score 4261; DB 8; Length 1506;

Best Local Similarity 34.1%, Fied: NO: 0  
Matches 819; Conservative 228; Mismatches

QY 2 QNSQDFSITELSLPKGGGAI TGMGEALTPTGPDGMAALSLPLISAGRGYAPFTLNYS 61

Db 3 QDSQDMVTQLSLPKGGAI SGMGDTISNAGPDGMASLSVPLPI SAGRGGAPNLSLNYSS 62

62 GAGNSPPGLGWDNCVMTIRRRTHFGVPHYDETDTFLGPEGEVLVA---DQP--RDEST 11

Db	63	GAGNGSFGIGMOSSTWAIISRRTOGHVPQYGHGDTFLCPMGVMAVAVNOSGQPDVRAKTDK	172
Qy	116	LOGINLGATFTVTVGYSRLESHPESRLLEYWQPKT-TGKTDFWLIYSPDGQVHLLGKSPQAR	174
Db	123	LLGGQLPVTVTVTRHQPRNIQHFSKULEYWPPTDVTETFWLWYSPDGGQIHFGEKTEQAQ	182
Qy	175	ISNPSQTTQTAOWLLSEASVSSRGEQIYYOYVRAEDDTGCEADEITHHLOATAQRYLHIVVY	234
Db	183	IANPAEVSQIAQWLLBEETVTPAGEHIYYOYVRAEDDTCDDSEKNAHPNASQAQRYLTQVNY	242
Qy	235	GNRTASSETPLGLDGSAPSQADWLFPYLVFDYGERSNMLKTPPAP-STTGSWLCQDPRPSRY	293
Db	243	GNITPSSLLVLKNTTPADNEWMLFHLVFDYGERAQEINTVPPPKAPSNMKIRPDRPSRF	302
Qy	294	EYGFERTRELRCQVLMYHHLOAL-DSKTEHNGPTLVSRLLIYNYDESALASTLVFVRRV	352
Db	303	EYGFVTRTRKCOQIIMPHFKLSLAGEIDGEBIPALVARKLUSYDLUNDSVTTILTAIRQM	362
Qy	353	GHEQDGNVTVLPLELAYQDFSPPRHHAHQMPDMVLANFNAIQRWQLVDLKGEGULGLLYQ	412
Db	363	AYETDATALPLEFDYQFPAKVTVQKQEMPQLAGLNAQFPVQLVDLYGEGISGLIYQ	422
Qy	413	DK-GAWYRSQAQRIGEIS-DAVTWKRMOPLSVIPSLQSNASLVNDINGDQLDWVITGPG	470
Db	423	DRPGAWMYQAPIRQKNVEDINAVTYSPIPLPKIPSQQDRATLMDIDGDLHDWVITAGAG	482
Qy	471	LRGVHSORPDGSTRPTPLNALPVEYTHPPAQLADLMGAGLSDLVLIGPKSVRLYXNTRD	530
Db	483	IQORYSNQPNGEWTHIPIISALPTEYFHPQADLVGAGLSDLALIGPRSVRLYANDRG	542
Qy	531	GFAGKXDVQSGDITLPPVPCADPRKLVAFSDVLGSGQAHLEVSATKVTWCPNLRGGRFG	590
Db	543	NWKAGINVPMPDGVNLPIPGCDASSLVASDMLGSGQQLHVEIAQSVKWCNPNLHGGRFG	602
Qy	591	QPTLPGFGOPATFNPAAQVYLADLPGSGPTDLIYVHTNRLDILPANKSGNGFPAEPVTLRF	650
Db	603	AAILLPGFSQPNCTFNANQVFLADIDSGTGADIIYAHSTVLDIYLNESGNGRFSAPVRLNL	662
Qy	651	PEGLRPDHTCOLQWADVQGLGVASLLISVPHSPHWRCDLTNWKMKPLNEMNNMGVHH	710
Db	663	PEGVMFNTCOLQVSDIQGLGAASIVLTVPHMTPRHRWYDFTHNKPWLLNVIINNRAET	722
Qy	711	TLVRSSSQFWLDBEKAALLTGOTPCYLPFPPIHTLWOTETDEISGNKLVTLTLRYARGA	770
Db	723	TLFVRSQAQFWLDBEKSQIELGFEASYLEFPPIHLLWRNEALDIBITGNRLTKVMNYAHGA	782
Qy	771	WDGRERFRGFGYVEQTDSHOLAQGNAPERT-----PALTKWYATGLPVIDNALSTEY	825
Db	783	WDGREREFGRVGTQIDTDEFAGKTT-EKAPDENIYPSRSISWFAATGLPEVDSQLPAEY	841
Qy	826	WR-DDQAQFAGFSRPTTW-----QDNKDVLPTEDDNSRYWPNRALKQLLSBELYGLD	878
Db	842	WRGDDQAFAGTFRFTRYEKGNAQEGQDTPIKEPTETEYWLNRAMKGOLLSEVYG-D	900
Qy	879	DSTNK-HVPVTVTEFRSOVERLQHTDSRYPVLWSSVVERSNVHYERTASDPQCSONTLS	937
Db	901	DKTEKAPIVTVTEARQVRLIPSNDBAAPSWSITIIENRSYHYERIVDPSCQVVLK	960
Qy	938	SDRFGQPLKLSQVPRQROPAINLYPDTLFDLKLANSYDDQOQLRLTYQQSSWHHLTN	997
Db	961	ADEYGPFLAKVDIAYPRENPKAPNPYDLSLPTDLPAUSDYDDQKQLVLTQQQSYVHLTQ	1020
Qy	998	NTVRVLGLPDSTRSDIFTY-----GAENVPAGGLNLELLSDKNSLIADDKPRYILGQOKTA	1053
Db	1021	QDDWVLGLTDSRYSEVVHYAQTDAQSDIPKAGLILEDLLKVDGLIGKDKFTFVLGQORVA	1080
Qy	1054	YTDQNTTLPQTPTRQALIAFTETTFNQSTLSAFNGSIPSDKLSLTLEAGYQOTNYLP	1113
Db	1081	YVGGD-----AEKPTRQVRVAYTETAAPFDDNALHAFDGVIAPELUTQOLLAGGY-----LLV	1132
Qy	1114	PRTGE-----DKVWVAHHGYTDYGTAAQFWRPQKQSNTOITGKITLTDWANYCVVQVTRD	1168

Db 1133 PQSDVAGSSKVVWVARQGYTEYSAQFYRPLIQRKSLLTGKVTLSHDTHYCVVVKTED 1192

Qy 1169 AAGTTTSKDYRFLTPVQLTDINDNQHLLTDLALGRPITLRFMTGKMTGYSSPEKA 1228

Db 1193 GAGMTTQAKYRFLPQALTINDNQHIVTFNALGQVTSRFGTNGKISGYSTPESK 1252

Qy 1229 SFSPPSDVNAIELKKPLPVAQCQVYAPESMPVLSQTFN---RLABQDQKLYNARI 1285

Db 1253 PFTVPDTEKALALQFTIPVSCNIIYVPSWMLLPQOGLTGQKKEGETLNNALHRAQW 1312

Qy 1286 TEDGRICITLAVRRVWQSKAIPQ---LISLNNGPRLPPHSLTLTDDRYDHDPEQOIRQ 1341

Db 1313 TEDGLICELAVRMKIKRQATSMMAVTLQOILAQTPROPFHAMITTDYDSDSQOQLRQ 1372

Qy 1342 QVPSDGFGRLLQAAARHAGVQRNEDGSLIIN-----VQHTENRWAVTGRTEYDNK 1395

Db 1373 SIVLSDGFGRLVQSAQRHAGEAWQRAEDGSLVVDNTGKPVVANTTTTRWAVSGRTEYDGK 1432

Qy 1396 GQPIRTQPYPLNDWRVYNSDSARQEKAYADTHVYDPIGREIKVITAKGFRITLTPW 1455

Db 1433 GQAIRAVLPYYLNDWRVYVSDSDSARD--DLVADTHFYDPLGREYQVKTAKGFRENMFMPW 1490

Qy 1456 FTVNEDENDTAAEV 1469

Db 1491 FVNVNEDENDTAAEL 1504

RESULT 13

AEBA47812

ID AEB47812 standard; protein; 1506 AA.

XX AC AEB47812;

XX DT 22-SEP-2005 (first entry)

XX DE Native XptB1(xb).

XX XptB1; toxin; insect resistance; insecticide.

XX Os Xenorhabdus bovienii.

XX US2005155104-A1.

XX PD 14-JUL-2005.

XX PF 23-DEC-2004; 2004US-00020848.

XX PR 07-JAN-2004; 2004US-0534893P.

XX PA (APEL/) APEL-BIRKHOED P C.

PA (HEYT/) HEY T D.

PA (THOM/) THOMPSON R L.

PA (MEAD/) MEADE T.

PA (LIZS/) LI Z S.

PA (RUSS/) RUSSELL S M.

PA (SHEE/) SHEETS J J.

PA (LIRA/) LIRA J M.

PA (FENC/) FENCIL K J.

PA (MITC/) MITCHELL J C.

XX

Pi Apel-Birkhold PC, Hey TD, Thompson RL, Meade T, Li ZS;

Pi Russell SM, Sheets JJ, Lira JM, Fencil KJ, Mitchell JC;

XX WPI; 2005-496874/50.

DR N-PSDB; AEB47811, AEB47817, AEB47819.

XX

PT New isolated protein and encoding nucleic acid having toxin activity

PT against an insect, useful for developing new insecticidal toxins that can

PT be used to control insects.

XX

PS Claim 3; SEQ ID NO 2; 54pp; English.

XX

CC This sequence represents native XptB1(xb). XptB1 is one of five TC (toxin

CC complex) proteins from Xenorhabdus. XptA1 is a "stand alone" toxin. XptA2

CC also has some stand alone toxin activity. XptB1 and XptC1 are the

CC Xenorhabdus potentators that can enhance the activity of either

CC both) of the XptA toxins. XptD1 has some level of homology with TcGB

CC (toxin complex c protein B). This protein has toxin activity against an

CC insect. The polynucleotide encoding this protein may be used to transform

CC a plant cell, and thereby generate a transgenic plant which may be

CC ingested by an insect, whose numbers are subsequently controlled. The

CC methods and compositions of the present invention are useful for

CC developing new insecticidal toxins and other proteins that can be used to

CC control insects.

XX

SQ Sequence 1506 AA;

Query Match 53.9%; Score 4261; DB 9; Length 1506;

Best Local Similarity 54.1%; Pred. No. 0;

Matches 819; Conservative 228; Mismatches 409; Indels 58; Gaps 21;

Qy 2 QNSQDSITHELSPKGGGATGCEALTPTGPGMAALSPLPISAGRGVAPATLNVNS 61

Db 3 QDSQDMVTVTQLSLPKGGGATGCGMDTISNAGPDGMAASLVPLPISAGRGGAPNLSLNVSS 62

Qy 62 GAGNSPFGLGWDCNVMTIRRTTHFGVPHYDDETDFLQPEGEVLVA---DQP--RDEST 115

Db 63 GAGNSGFGIGHQSGSTMAISRTQHGVPQYQGEDTFLCPMGEVMAVANQSQQPDVRKTDK 122

Qy 116 LQGINLGATFTVTGYSRSLRSHFSRLEYWQPKT--TGKTDFWLIYSPDQVHLLGKSPQAR 174

Db 123 LGGQLPVTVTTRHQPNRIQHFSLKLEYWQPTDVTETPFWLMYSPDQIHIFGKTEQAQ 182

Qy 175 ISNPSQTTQTAQMLLEASVSRGEQIYYQRAEDDTCGEADEITHHQAQRYLHIVYY 234

Db 183 IANPAEVSQIAQWLLLEETVTPAGEHIYYQYRAEDDTCGDDSEKNAHPNASEAQRYLTQVNY 242

Qy 235 GNTTASETLPLDGSAPSQADWLFYLVFDYGERNNLKTTPAF--STTGSMLCRODRSRY 293

Db 243 GNITPESSLVLKNTVPADNEWLFHVLVDYGERAQEINTVPPFKAPSNKIRPDRSRF 302

Qy 294 EYGFETRRLCQVLMYHHLQAL--DSKITENHGTFLVSRLLILNYDESASTLTVFVRV 352

Db 303 EYGFETRRLCQVLMYHHLQAL--DSKITENHGTFLVSRLLILNYDESASTLTVFVRV 362

Qy 353 GHEQDGNVTLPLLELAYQDFSPRHAWQPMQVLANFNALQRMQLVDLKEGELPGLLYQ 412

Db 363 AYETDATALIAPLEFDYQPFQAKVTKQWQEMPQAGLNAQQPYQLVDLYGEGISGLYQ 422

Qy 413 DK--GAWVRSQAQRLGEIGS--DAVTWKMQLSVLPISQSNASLVINDGOLDWVITGPG 470

Db 423 DRPGAWVYQAPIROKVNEDINAVTYSFINLPKIPQDDRATLMDIDGDLHDMVIAGAG 482

Qy 471 LRGYHSQRPDGSWTRFTPLNALPVEYTHPRAQLADLMGAGLSDLVLIGPKSVRLYANTRD 530

Db 483 IQGRYSNQPNGEWHFTFIPISALPTEYFHPQALADLVAGLSDLALIGPSRVLYANDRG 542

Qy 531 GFAKGDVQSGDITLVPQADPRKLVAFSVIGSGQAHLVEVSATKVTCPNIGRGRFG 590

Db 543 NWKAGINVMPPDGVNLFIFGCDASSLVAFSDMLSGQQHLVEIAAQSVCWPNLGHGRFG 602

Qy 591 QPITLPGFSOPATEFNPAQVYVLADLQSGPTDLYVHTNRDLDFLNSKNGFASFPVTLRF 650

Db 603 AAILLPGFSQPNGFNFANQVFLADIGSGTADIIYAHSTDLIDYLNESGNRFSAPVRLNL 662

Qy 651 PEGRLFDHTQLOQADVQGLGVASLILSVPHMSPHWRCDLTNNKPLNLMNNMNMVHH 710

Db 663 PEGVWFNTQQLQVSDIQGLGAASIVLTVPHMTPRHRYDFTHNKPILLNINNRRGNET 722

Qy 711 TLRYRSSQFWLDEKAAALTTGQPPVCYLPPIHTLMQTEDETSIGNKLVTLTRYARGA 770

Db 723 TLFYRSSAQFWLDEKSKQIEELGKFAASYLPFPIHLLWRNEALDEITGNRLTKVMVYAHGA 782

Qy 771 WDGEREPRGEGYVEQDTHQLAQGNAPERTP-----PALTKWNYATGLPVIDNALSTEY 825

Db 783 WDGEREPRGEGYVEQDTHQLAQGNAPERTP-----PALTKWNYATGLPVIDNALSTEY 841



```
Qy 713 YRSSQFWLDEKAAALTTGTPVCYLPPFPIHTLMQTEDEISGNKLVTLRYARGAWD 772
Dd ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Dd 720 HYRSSVQFWLDEKAAALAAAGSPACYPFTLTLTLWRSVVQDEITGNRLVSDVLYRHGWD 779
Qy 773 GREERFGVGVETQDSHOLA-QGNAPERTPALTKNMYATGLPVIDNALSTEWYRDD-Q 830
Dd ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Dd 780 GQERFGRFGFVETRTDTTLASQGTATLSLSPSVSRNWTATGVPADVERPETWQNDAA 839
Qy 831 AFAGFSRFTTWQDNKDVPLTPEDDNRNFWNRALKGQLLRSELYGLDDSTNKHVPYTVT 890
Dd ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Dd 840 AFADPATFTVSGSEDEQYTP-DDSKTFWLQRLKGLLSELYGADGSSQADIPISVT 898
Qy 891 EFRSQVRLQHTDSRYVPLWMSVSVESNHYHYERIASDPQCSQNTLSSDRPGQLKLSV 950
Dd ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Dd 899 ESRPQV-RLVEANGDYPVWPMGAESRTSVYRYHNDPQCOQAVALLSDEYGFPLRQVSV 957
Qy 951 QYPRRQPAINLYPDTLPDKLLANSYDDQQLRLTYQSSWHHTNTNVR--VLGLPDS 1008
Dd ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Dd 958 NYPRRPSADNPYPASLPATLAFANSYDEQQOILRLGLQQSSAHLVLSLSEGHVLLGLAE 1017
Qy 1009 TRSDIFYGAENVAGGINLELLSDKNSLIADDPREVLGQOKTAYTDGQNTTPLQTPTR 1068
Dd ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Dd 1018 SRDQVFTYADNVPEGGLLLEHLLAPESLVSQVGLTAGQQVWYLDSDQVATVAAPPL 1077
Qy 1069 QALIAFTETTVFNOSTLSAFNGSIPSKLSTLEQAGYQNTNLYFP--RTGEDKVVVAHH 1126
Dd ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Dd 1078 PPKVAFIETAVLDGVMVSSLAAYIVDEH---LBQAGYRQSGYLFPRGREAEQALWTQCQ 1133
Qy 1127 GYTDYGTAAQFWRPQKQSNQTLTKGKILIMDANYCVVVQTRDAAGLTSKAYDWRFLTPV 1186
Dd ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Dd 1134 GYVYAGAEHFWLPLSRDMSMLTGFVITRDAYDCVITQWQDAAGIVTTADYDWRFLTPV 1193
Qy 1187 QLTIDINDQHLITDALGRPITLTFWGTENGKMTGYSPKASPPSDVNAALTELKPL 1246
Dd ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Dd 1194 RVTDPNDNLQSVTLDALGRVTLTFWGTENGIAIGYSD--ATLSVPDGAALALATPL 1250
Qy 1247 PVAOCQVAPESMMPVLQSKTFNRLABQDMOKLNNARIITEDGRICTLAYRRVQSQAI 1306
Dd ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Dd 1251 PVAQCLVYVTDW-----GDDNDEK----- 1270
Qy 1307 POLLSLNNGRPLPHSLTLTDRYDHPDEQOIRQOVVFSDFGRLLOAAARHEAGMARQ 1366
Dd ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Dd 1271 -----MPHVVVVLTADRYSDTGQOVRQOVTFSDGFGRELQSATROEGNAWQ 1318
Qy 1367 RNEDGSLI-----INVQHTENRWAVTGRTEYDNKGOPIRTYQYFLNDWRVYNSDSAR 1419
Dd ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Dd 1319 RGRDGLVTASDGLPVTVV-TNFRWATVGRAYDNKGLPVRYVQYFLDSHQYVSDDSAR 1377
Qy 1420 QKEBAYADTHYVDPIGREIKVITAKGWRPRTLFTFWFTVNEDENDT 1465
Dd ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Dd 1378 Q--DLVADTHFYDPTAREWQVITAKGERRQVLYTFWFWVSEDENDT 1421

RESULT 15
ADR21540
ID ADR21540 standard; protein; 1428 AA.
XX
AC ADR21540;
XX
DT 04-NOV-2004 (first entry)
XX
DE Serratia SepB toxin SEQ ID NO:60.
XX
KW toxin; insect; insecticidal; transgenic; pest control.
XX
OS Serratia entomophila.
XX
PN WO2004067727-A2.
XX
PD 12-AUG-2004.
XX
PF 07-JAN-2004; 2004WO-US000394.
```

```
XX 21-JAN-2003; 2003US-0441723P.
XX (DOWC ) DOW AGROSCIENCES LLC.
XX
XX Hey TP, Schleper AD, Bevan SA, Bintrim SB, Mitchell JC, Li ZS;
XX Ni W, Zhu B, Merlo DJ, Apel-Birxhold PC;
XX WPI; 2004-580999/56.
XX
XX Controlling or inhibiting an insect, useful for pest control, comprises
XX contacting the insect with effective amounts of a Protein A, a Protein B,
XX and a Protein C.
XX
XX Claim 1; SEQ ID NO 60; 368pp; English.
XX
XX The invention relates to a novel method for controlling or inhibiting an
XX insect comprising contacting the insect with effective amounts of a
XX Protein A, a Protein B, and a Protein C. Each of the Proteins A, B, and C
XX is encoded by a naturally occurring gene or has an amino acid sequence
XX that differs from the product encoded by a naturally occurring gene only
XX by truncation or by conservative amino acid changes. Protein A is a 230-
XX 230 kDa toxin complex insect toxin that is derived from a first taxonomic
XX species, has stand alone insecticidal activity, and has an amino acid
XX sequence at least 40% identical to a sequence selected from XptAlwI,
XX XptA2wI, TcdA1, TcdA2, TcdA4, and TcdA. Protein B is a 130-180 kDa toxin
XX complex potentiator having an amino acid sequence at least 40% identical
XX to a sequence selected from TcdB1, TcdB2, TcdC, XptC1wI, XptB1xb,
XX PtcB1(orf5), or SepB. Protein C is a 90-120 kDa toxin complex potentiator
XX having an amino acid sequence at least 35% identical to a sequence
XX selected from TccC1, TccC2, TccC3, TccC4, TccC5, XptC1xb, XptC1xb, PtcC1
XX (orf 6 long), PtcC1 (orf 6 short), and SepC. Also claimed is a transgenic
XX plant or plant cell that produces a Protein A, a Protein B, and a Protein
XX C. The method is useful for pest control. The present sequence represents
XX Serratia entomophila SepB toxin.
XX
XX Query Match 50.3%; Score 3971; DB 8; Length 1428;
XX Best Local Similarity 52.7%; Pred. No. 2.3e-309;
XX Matches 783; Conservative 186; Mismatches 431; Indels 86; Gaps 16;
Qy 1 MONSDPSITELSLPKGGGAICTMGREALPTGPDGMAALSPLPISAGRYAPAFITLYN 60
Dd ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Dd 1 MONHQDMAITAPTLPSCGGAVTGLKIDAAAGDGAATLSIPLVSPGRGYAPATGALNYH 60
Qy 61 SGAGNSPPGLGWCNVTIRRTTHFGVPHYDEDTFLGPEGEVLVA-----DQPRDES 114
Dd ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Dd 61 SRSGNGFGGIGGGAAGVQRRTRNGAPTVDYDDTDFGPDGVLVPAUTAAQTQEARQAT 120
Qy 115 TLQGINLGATFTVTGYSRSLRSHFSRLVYWPQTKTKTDFWLIYSPDGQVHLLKSPQAR 174
Dd ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Dd 121 SLGGINPGSFNVQVRSRTEGSLRSLRLWLPADETETEFWLYTPDQVALLGRNAQAR 180
Qy 175 ISNPSOTTQPAQWLLSEASVSSRGEIYQYVRAEDDTGCEADEITHHLQATQARYLHYVY 234
Dd ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Dd 181 ISNPTAPTQAVWLMESSVSLTGEQWYQYVRAEDDDGCEADAHQPAQARYPVAVMY 240
Qy 235 GNRITASETLPLDGSAPSQADWLFLVFDYGERSNMLKTPPAPSTTGS--WLCHQDRFSR 292
Dd ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Dd 241 GNRQAARTLPAL--VSTPMSDMSWLFILFDYGERSSVLSSEAPAWQTPGSGEMLCRQDCFSG 299
Qy 293 YEYGFETRRLRCQVLMVHHQLALDSKITEHNGPTLVSRLLIYNDSASATSLVFFVRV 352
Dd ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Dd 300 YEFQFNLRRLRCQVLMFHYLVLAGSSGANDAPALISRLLLDYRSPSLLENVHQV 359
Qy 353 GHEQDGNVVTLPPLLAYQDFSPRRHAHWQMDVLANFNANAIQRWLQVLDKGEGLPLLYQ 412
Dd ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Dd 360 AYESDGTSCALPALALGWQTFPTTSLAWQTRDDMGKLSLLQPVQLVDLNGEGVVGLYQ 419
Qy 413 DKGAWYTRSQRLEIGSDAVTWKMQPLSVIISLQSNASLVINDINGQGLDWTITGGLR 472
Dd ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Dd 420 DSGAWYREFVROSGDDPDVATWGAAAALPTMPALHNSGILADLNGDGRLEWVVVTAFCVA 479
```





GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 16, 2005, 14:33:11 ; Search time 54 Seconds  
(without alignments)

2626.361 Million cell updates/sec

Title: US-10-706-424-10

Perfect score: 7901

Sequence: 1 MQNSQPSITSLPKGGGA.....WFTVNEDENTAAEVKKVKM 1474

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR.80.\*

2: PIR1.\*

3: PIR2.\*

4: PIR3.\*

5: PIR4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3569.5	45.2	1496	2 AH0447	insecticidal toxin
2	839	10.6	591	1 S09498	virulence-associat
3	835	10.6	591	1 S22664	virulence-associat
4	822.5	10.4	593	2 S15215	virulence-associat
5	709.5	9.0	563	2 S78224	virulence-associat
6	172.5	2.2	1404	2 E85509	hypothetical prote
7	172.5	2.2	1404	2 E30658	RhEg core protein
8	164.5	2.1	1399	2 A99720	RhEg core protein
9	161	2.0	1863	2 S46217	protein-tyrosine-p
10	160.5	2.0	1426	2 H4780	RhAd protein precu
11	158	2.0	3083	2 A12433	hypothetical prote
12	156	2.0	2167	2 A1499	cell wall-associat
13	155.5	2.0	1400	2 E30886	RhEg core protein
14	153	1.9	3036	2 T18995	hypothetical prote
15	151	1.9	1377	2 E86034	RhA protein in rh
16	151	1.9	1394	2 H91236	RhA core protein
17	151	1.9	1409	2 F91187	RhA core protein
18	151	1.9	2314	2 T28698	hypothetical prote
19	149	1.9	1397	2 A85570	RhC protein in rh
20	149	1.9	1398	2 B85549	hypothetical prote
21	149	1.9	1907	2 S50893	protein-tyrosine-p
22	148	1.9	1398	2 H30698	RhAd core protein
23	148	1.9	3972	2 S75251	hypothetical prote
24	147.5	1.9	682	2 C64898	RhA protein - Bsc
25	147.5	1.9	4199	2 S76412	hypothetical prote
26	145.5	1.8	632	2 P83387	copper resistance
27	144	1.8	709	2 S38241	hypothetical prote
28	144	1.8	1377	2 C65159	RhA protein precu
29	144	1.8	1411	2 E65145	RhB protein precu

30	142.5	1.8	1512	2 AH0439	probable membrane
31	141	1.8	1274	2 T37193	enamelin matrix pr
32	141	1.8	1397	2 C64805	RhC protein precu
33	141	1.8	2893	2 A64556	toxin-like outer m
34	139.5	1.8	1577	2 T30858	glucosyltransferas
35	139	1.8	828	2 G87584	hypothetical prote
36	139	1.8	2515	2 S47008	tenascin-like prot
37	138.5	1.8	1158	2 F90854	probable host spec
38	138	1.7	4427	2 FN0637	polyketide synthas
39	137.5	1.7	656	2 G85731	RhA element associ
40	137	1.7	2514	2 F81045	hemagglutinin/hemo
41	137	1.7	2628	2 T28651	hemagglutinin A -
42	136.5	1.7	1645	2 H85554	hypothetical prote
43	136.5	1.7	3566	1 A40701	tenascin-X precurs
44	136	1.7	775	2 C95921	hypothetical prote
45	136	1.7	1083	2 T23031	hypothetical prote

#### ALIGNMENTS

##### RESULT 1

AH0447

insecticidal toxin complex [imported] - Yersinia pestis (strain CO92)

C;Species: Yersinia pestis

C;Date: 02-Nov-2001 #sequence\_revision 02-Nov-2001 #text\_change 09-Jul-2004

C;Accession: AH0447

R;Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Hoiden, M.T.G.; Prentice, M.B.

deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;

il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell,

Nature 413, 523-527, 2001

A;Title: Genome sequence of Yersinia pestis, the causative agent of plague.

A;Reference number: AB0001; MUID:21470413; PMID:11586360

A;Accession: AH0447

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-1496 <KUR>

A;Cross-references: UNIPROT:Q8ZAV4; UNIPARC:UPI000000CDA67; GB:AL590842; PIDN:CAC93148.1;

C;Genetics:

A;Gene: YPO3678

Query Match	45.2%	Score	3569.5;	DB 2;	Length	1496;			
Best Local Similarity	49.3%	Pred. No.	3.8e-220;						
Matches	748;	Conservative	211;	Mismatches	486;	Indels	71;	Gaps	29;
Qy	1	MONS-QDPSITSLPKGGGAI	TGMBALPTGPDGMAALSLPLPISAGRGYAPAF	TNY	59				
Db	1	MENSKQAVAPLSLPKGGGAI	TGMBALPTGPDGMAALSLPLPISAGRGYAPAF	LSY	60				
Qy	60	NSGAGNSPFLGMDCNVMT	IRRTTHFGVPHYDETDFTLGPGEVYLWVADQPRD	--ESTLQ	117				
Db	61	SSGSGNGPFLGMDCNVMT	IRRTTHFGVPHYDETDFTLGPGEVYLWVADQPRD	--ESTLQ	120				
Qy	118	GINLGATFTTGYRSRLES	HFSLRYWQPKT-TGKTDFFWLIYSDGQVHLKGFQARIS	176					
Db	121	SLN-GEQFVIRYLPRIE	GNFRHRYEYRPTNNSQAPFVHVSDDGQVHLKGFQARIS	179					
Qy	177	NPSQTTQAWLLASVSSR	GEQIYYQVRAEDDTGCDADET-----THHLQATQRYLHI	231					
Db	180	DPLPHEHIAEWLLAESV	LSGHEHGYQQAEDGIDEPSIYKAEKQNHPPAASQRYLKR	239					
Qy	232	VYGNRTTASETLPGD	GSAPSQADWLFYLPVDPYDGERNNLKTTPAFSTGTSWLCRQDRFS	291					
Db	240	VYGNRQAAYELCLT-Q	QAPATSWLFLIFDHGYESNIAEQVEPVIYKGSWNRQDAFS	298					
Qy	292	RYEGFPIRTRLCRQV	LMTHLQALDSKITEHNGPTLVSLRLINLYDESATSLTVFVRR	351					
Db	299	HFNYGFVTRRLCQV	LMTHLQALDSKITEHNGPTLVSLRLINLYDESATSLTVFVRR	357					
Qy	352	VGHEQDGNVVTLP	PLELAYQDFSRRHAHQPMQDVLANFNAIORWQLVDLKGEGPLGLY	411					
Db	358	LAHPDGTGKSLP	PLEFEDYQDFSTRDALGWLQPLTDWAEFN--YQYQWDLNGEGMPGLY	415					





A:Molecule type: DNA  
A:Residues: 1-593 <XRA>  
A:Cross-references: UNIPROT:P24419; UNIPARC:UPI00001780F6; EMBL:X56727; NID:g47836; PIDN:  
A:Experimental source: pEX102  
A:Note: the authors translated the codon CTG for residues 54 and 56 as Val  
R:Taira, S.; Baumann, M.; Riikonen, P.; Sukupolvi, S.; Rhen, M.  
FEMS Microbiol. Lett. 77, 319-324, 1991  
A:Title: Amino-terminal sequence analysis of four plasmid-encoded virulence-associated p  
A:Reference number: A54540  
A:Accession: A54540  
A:Status: preliminary  
A:Molecule type: protein  
A:Residues: 1-10 <TA1>  
A:Cross-references: UNIPARC:UPI00001780F7  
R:Taira, S.; Rhen, M.  
Microb. Pathog. 7, 165-173, 1989  
A:Title: Identification and genetic analysis of mkaA-a gene of the *Salmonella typhimurium*  
A:Reference number: S26565; MUID:90136009; PMID:2893884  
A:Accession: S26565  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-56, 'P', 58-304, 'L', 306-340, 'R', 342-373, 376-492, 'V', 494-593 <TA2>  
A:Cross-references: UNIPARC:UPI00001780F8; EMBL:Z15042; NID:g47782; PIDN:CAA78760.1; PIDN:  
A:Note: the authors translated the codon GTT for residue 493 as Gly  
C:Genetics:  
A:Gene: vscC; mkaA  
A:Genome: plasmid  
C:Superfamily: virulence-associated protein spvB  
C:Keywords: DNA binding; transcription regulation

	Query Match	10.4%	Score 822.5;	DB 2;	Length 593;
	Best Local Similarity	47.3%;	Pred. No. 1.1e-44;		
	Matches 175;	Conservative 52;	Mismatches 116;	Indels 27;	Gaps 8;
Qy	9	ITELSLPKGGATGCEALTPTGPDGMAALSLPLISACRGYAPAFLLNYSAGNSPF	68		
Db	15	ITPPFLPKG-----GKALSQSGPDGLASITLPLPISAEGRFAPALALHYSGGNGPFP	67		
Qy	69	GLGWDGNVMTIRRTTHFGVPHYDETDTFLGPEGEVLV----VADQPRDSETLQ--GINLG	122		
Db	68	GVGWSNCATSIARTSHGVQPYNDSDFLPGDGEVLVQLSTGDAENPVTCFAYGDVSPF	127		
Qy	123	ATTFTVGYRSLRSHFSRLEYWPKTKGTFDWLIYS PDGQVHLGKSPQARISNPSQTT	182		
Db	128	QSVTVTRYQPTRESSFVRLRYWVGNNG--DDFWLLHDSNGILLHLGKTAARLSDPQAA	186		
Qy	183	QTAQWLLEASVSSRGHQIYYQYPAEDDTGCEADEITHLQATAQRYLHVIYVGNRTASET	242		
Db	187	HTAQWLVEESVTPAGEHIYYSYLAENGDNVDLNGNEAGRDRSAMRYSKVKYQGNATPAAD	246		
Qy	243	LPGLDGSAPQADWLFLVFDYGERSNLTKTPAPFSTTGSWLCQDRFSRYEYGFELRTR	302		
Db	247	L-YLWTSATPAVQWLFTLVFEYGERGVDPQVFPAPFAQAQNSWLARQDPFSLYNTGFELRLH	305		
Qy	303	RLCRQVLMVHHLQALDLSKTHNGPTLVSLRLIINYDESAITSLVFRVRRVGHQBQD----	358		
Db	306	RLCRQVLMVPH----FPDELGE--ADTLVSRLLLEYDENPILTOLCAARTLAYEGDGYRRA	360		
Qy	359	----NVVTLPP	365		
Db	361	PVANNMPPPP	370		

RESULT 5  
S78224  
virulence-associated protein mkbB - *Salmonella typhimurium* plasmid  
C:Species: *Salmonella typhimurium*  
C:Date: 29-Jan-1998 #sequence revision 13-Feb-1998 #text\_change 20-Aug-1999  
C:Accession: S78224; S23713; A43996  
R:Norel, F.  
submitted to the EMBL Data Library, January 1992  
A:Reference number: S78224  
A:Accession: S78224

A;Molecule type: DNA  
A;Residues: 1-563 <NOR>  
A;Cross-references: UNIPARC:UPI000008BA96; EMBL:X57096; NID:g46998; PIDN:CAA40380.1; PID  
R;Norel, F.; Pisano, M.R.; Nicoli, J.; Popoff, M.Y.  
Res. Microbiol. 140, 455-457, 1989  
A;Title: Nucleotide sequence of the plasmid-borne virulence gene mkb from Salmonella ty  
A;Reference number: A43996; MUID:90161559; PMID:2696057  
A;Accession: 823713  
A;Molecule type: DNA  
A;Residues: 1-513, 'NILEKHKGRILGDVAHFKGEAEMLF', 514-563 <NOW>  
A;Cross-references: UNIPARC:UPI00001780F5; EMBL:X57096  
C;Genetics:  
A;Genome: plasmid  
C;Superfamily: virulence-associated protein spvA  
C;Keywords: DNA binding; transcription regulation

Query Match 9.0%; Score 709.5; DB 2; Length 563;  
Best Local Similarity 39.1%; Pred. No. 1.7e-37;  
Matches 170; Conservative 60; Mismatches 138; Indels 67; Gaps 13;

QY 9 ITLSLPKGGGATGMEALTPTGPGMAALSPLPISAGRYAPAPTLNYSAGNSPF 68  
DB 15 ITPPFLPKG-----GKALSQSGPDGLASITLPLISAERGAPA----- 53  
QY 69 GLGWDCNV-----MTIRRTTFG-----VPHYDETDTFLGPEGLV-----VAD 108  
DB 54 ---WRCTTAAVAAMALRGLVLRDNEHCPPQPCVPOYNDSDRFLGPDGEVLVQTSTGD 110  
QY 109 QPRDESTLQ--GINLGATFTVGYRSRLESFHRLEYWQPKTKTKTDFTWLIYSPDGOVHL 166  
DB 111 APNPVTSFAYGDVSFPQSYTVTRYQPTRESSFYRLEYWVGNNG--DDFWLLHDSNGILHL 169  
QY 167 LKSPQARISNPQTTQATQWLLEASVSSRGEIYQYRAEDDTGCEADRIITHLQNTAQ 226  
DB 170 LGKTAARLSDPQAASHTAQLWEESVTPAGEHIYVSYLAENGNDVNLNGNEAGDRSAM 229  
QY 227 RYLHIVYGNRTASETLPLGIDGAPSOADMFLVFDYDGERSNLKTTPPAFTTGSMLCR 286  
DB 230 RYLSKVQYGNATPAADL-YLWTSATPAVQWLFVLVDYGERGVDPQVPAPTAQNSWLAR 288  
QY 287 QDRFSRYEYFEINTRRLCRQLVMYHLQALDCKITEHNGPTLVSRILNLYDSATASTL 346  
DB 289 QDPFSLYNGFEIRLHRLCRQLVMFHH---FPDELGE--ADTLVSRLLLEYDENPILTQL 343  
QY 347 VFVRVCHQDG-----NVVTLPLLEYAQDFSPRHHAWQPMQMDVLANFNAI-----Q 394  
DB 344 CAARTLAYEGDYRRAPVNNMPPPPPPMMGGNSRSPKSKWAIVESKQIQALRYYSQAQ 403  
QY 395 RWQLVD--LKGEGLP 407  
DB 404 GYSVINKYLRGDDYP 418

RESULT 6  
E85509  
hypothetical protein Z0268 [imported] - Escherichia coli (strain O157:H7, substrain EDL9  
C;Species: Escherichia coli  
C;Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 09-Jul-2004  
C;Accession: E85509  
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew  
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimantanta, E.; Potamouasis, K.; Apodaca,  
Nature 409, 529-533, 2001  
A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
A;Reference number: A85480; MUID:21074935; PMID:11206551  
A;Accession: E85509  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-1404 <STO>  
A;Cross-references: UNIPROT:Q8XED9; UNIPARC:UPI00000D025B; GB:AE005174; NID:g12512977; F  
A;Experimental source: strain O157:H7, substrain EDL933  
C;Genetics:  
A;Gene: Z0268  
C;Superfamily: rhes protein

Query Match 2.2%; Score 172.5; DB 2; Length 1404;  
Best Local Similarity 19.4%; Pred. No. 0.021;  
Matches 254; Conservative 137; Mismatches 427; Indels 489; Gaps 69;

QY 327 PTIVSRILIYNDESATIASTLVFVRRVGHQDGNVVTLPPLLEYAQDFSPRHH----- 379  
DB 51 PVLGAKVLPGETDLALPGPLPFI-----LSRAYSSYRTTTPAPVGVFGP 94  
QY 380 HQQ-PMDVLANFNAIQRWQLVDLKGEGPLGLLYQDKGAWWYRSARQLGEIGSDAVTWK 438  
DB 95 GWKAPFDI-----RLQIRD-----EGLILNDN-----GGRSIHPEPL 126  
QY 439 QPLSVIPSLQSNASLVDINGDQLDQWVTGPGRLGVHSORP-DGSVTRFTPLNALPVEY- 496  
DB 127 PPGEI--SYRSSEF-----WLARG-GVAEOHSSQPLSALW-----QVLPEVDR 167  
QY 497 THPRAQLA--DLMG-----AGLSDLVLIGPKSVRLYANTRDGF-----AK 534  
DB 168 LSPHMYLATNSLQGPWMLNWPPEVFGADEVLPPPEPAYRVLTVGVVDGFGRTLAFHAAE 227  
QY 535 GKDVVQSGDITLVPQADPKLVAFSDVLGSGQAHLEV-----SATKVT--CHPNLGRGR 588  
DB 228 G-DV--AGAVTGVTDGAGRR---FHLVL-TTQQAQAEVFRKQATSLSSPAGFRSASS 279  
QY 589 FGQPIITLPGFSQPATEF-----NPAQVYLADLDGSGPTDLIYVHTNRLDI 633  
DB 280 LVFPDTPA-----GTGYADNGIRLEAVWLTHDPA--YPDELPAAPLARYTYTASSELRA 333  
QY 634 FLNKGSG--NGFAPBVTLPFPPEGLRFDHTCQLOMADVOGLGVASLILSVPHMSPHH--- 686  
DB 334 VYDRSGTVRGFA-----YDAEHAG-----RMVAHVAGR 363  
QY 687 ---WRCDLTNMPKPLNEMNNMNGVHHTLYRYSQFQWLDKAAALTTGQTPVCYLPPF 742  
DB 364 PESRYRYDDTGR---VTELNVNPEGLDYFEGQ-----DRVTITDSLNRREVLY--- 409  
QY 743 IHTLWQTEDEISGNKLVLTTLRYARGAMDGRERFRFGYVEQTP-----SHQLAQ 794  
DB 410 -----TEGE--GGLKRVVKEHADGSTRSEYDEAG-RLKAQTDAAGRRTYSLHMAS 459  
QY 795 GNAPERTPP-----ALTKNMYATGLFVIDNALSTEYMRDDQAFAGSPRFTT 841  
DB 460 GAVTAVTGPDRTVRYGYNQROVTSVTPDGL-----RSSREY----- 498  
QY 842 WQNDKQVLTPEDDNSRYWFRNALKQLLRSELYGLDDSTNKIVPTVTEFRSQVRLOH 901  
DB 499 -----DEKGLAAETSRSGETTR---YSYDDPASE-LPTGIQD-----A 533  
QY 902 TDSRYPVLWSSVVESRNYHYERTASDPQCSNIT-LSSDRFGOLPKQLSVQYPRQCPAI 960  
DB 534 TGS TKQWMS-----RYQLLITTDGSGYTRYRYDYRQY---QIAV-----HREGI 578  
QY 961 NLYPDTLPKLLANSYDDQORQLRLTYQSSWMHLLTNNTVRLVGLPDPSTRSDI--FTYG- 1017  
DB 579 STYSYNPRQLVSQKDAQGRETR--YEYSAAGDLT-----AIVAPDGSSEIYQDAWGK 631  
QY 1018 AENVPAAGL-----NLELLSKNSLIADDK--PREYLCOQKTAITDQGN----- 1059  
DB 632 AVSTTQGLTRSMGYDAAGRITVLTNENGSGSTFRYPDPVDRLTEQRG--FDGRTQRYHYD 689  
QY 1060 -TTPLOTPTTQALIAFTETTTFVNQSTLSAPNGSIPSDKLSLTTLLEOAGYQOTNYL--PPRT 1116  
DB 690 LTGKLQSEDEGLVTLWHYDASDRITRVTNGD-----PAEQWYDEHGLWTLTSH 741  
QY 1117 GEDKVMVAHHGYTDYG-----TAAQFWRPQ-----KQSNQTLQTKG 1151  
DB 742 SEGRVSHVHYGYDDKGLRTGERTVENPETGEMLEWHEHTGHAYSEQGLATRQSPDGLPPV 801  
QY 1152 ITLIWDANYC-----VVQTRD-----AAGLTTSKAYDWRLTLP 1185  
DB 802 EWLTGSGYLAKMKGTLGTPLVYMRDLRHRETARSFGGEAYELATAMNTSGQLRSRLNL 861

```
Qy 1186 VOLT---DINDNQHLLTLDALGRPITLRFWGTENGKMTGYSSPEKASPPSDVNAIEL 1242
Db 862 POLDRDYDNDNGQLIRISGPQSEYRY--SDTGLRTGVHT-----TAANLDIDI 910
Qy 1243 KKPLPVAQCQVAPESWMPVLSQKTFNRLAQDQWQKLYNARIITEDGRICTLAYRRWQS 1302
Db 911 -----PYATDPAGNRLPD----- 923
Qy 1303 QKAIPQLISLLNNGPRLPHSLTLTTRDYDHPDQIQIRQ---VVFSDGFGRLLOAAARH 1359
Db 924 -----PELHPDS-TLTA-----WPNRRIAEADAHVYRYDEVYGRLEAKTDRI 963
Qy 1360 EAGMARQENEGSLIINVQHTENRWATGRTYDKNKQPIRTYQPYPLNDWRYVNSDAR 1419
Db 964 PEGVIRMHDE-----RTHHYHDSQHLVFHTRIQ 993
Qy 1420 QEKEAYADTHVDPIGRIKIVITAG-WFRRTLFTPMFTVNEDENDT 1465
Db 994 HGEQVESRYLYDPLGR-----TGKRVRRERDLTGWMSLSRKPEET 1036

RESULT 7
RhsG core protein with extension [imported] - Escherichia coli (strain O157:H7, substrain
C;Species: Escherichia coli
C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 09-Jul-2004
C;Accession: E90658
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and geno
A;Reference number: A99629; MUID:21156231; PMID:11258796
A;Accession: E90658
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1404 <HAY>
A;Cross-references: UNIPROT:Q8XED9; UNIPARC:UPI00001653A9; GB:BA000007; PIDN:BA033660.1;
A;Experimental source: strain O157:H7, substrain RMD 0509952
C;Genetics:
A;Gene: ECa0237
C;Superfamily: rhesF protein

Query Match 2.2%; Score 172.5; DB 2; Length 1404;
Best Local Similarity 19.4%; Pred. No. 0.021;
Matches 254; Conservative 137; Mismatches 427; Indels 489; Gaps 69;

Qy 327 PTLVSRLLNYDESAIASTLVFVRVGHQGVNVLPLLELAYODRSPRHA----- 379
Db 51 PVLGAKVLPGETDLALPGPLPFI-----LSRAYSSYKTRTPAPVGVFGP 94
Qy 380 HWQ-PMDVLANFNATQRMQLVDLKGEGPLGLLYQDKGAWYRSQRLGEIGSDAVTWKEM 438
Db 95 GWKAPFDI-----RLQIRD-----EGLILNDN-----GGRSIFHEPL 126
Qy 439 QPLSVIPSLQNASLVINGDQGLDWITGPLGRGYHSQRP-DGSWTRFTPLNALPVEY- 496
Db 127 PFGBI--SYRSSESP-----WLAGR-GVAEQHSSQPLSALW-----QVLPEDVR 167
Qy 497 THPRAQLA--DLMG-----AGLSDLVLIGPKSVRLYANTRGF-----AK 534
Db 168 LSPHMYLATNSLQGPWTLNHPVRVPGADVLEPPPEPAIRVLTGVVDGFGRTLAFHRAAE 227
Qy 535 GKDVVQSGDITLVPVGADPRKLVAFSDVLGSGQAHLEVE-----SATKVT--CWNLGRGR 588
Db 228 G-DV--AGAVTGVTDGARR-----FHLVL-TTQRAEVFRKQKQATSLSSPAGRSASS 279
Qy 589 FGQPIITLPGFSQPAIEF-----NPAQVYLADLQSGGPTDLIYVHNRLDI 633
Db 280 LVFPDITLPA-----GTEYGADNGIRLEAVWLTHDPA--YPDELPAAPLARYTYTASGELRA 333
Qy 634 FLNKG---NGFAPVTLRFPEGLRFDHTCQLQADVQGLGVASILLVSPHMSPHH---- 686
Db 334 VYDRSGTQVRGFA-----YDAEHAG-----RMVAHHYAGR 363
```

```
Qy 687 -----WRCDLTNMPKWLNNMNNNGVHHTLRYSSSQFVLDKAAALTTGQTPVCYLPFP 742
Db 364 PESRYRYDDTGR-----VTELVNPEGLDYREYQG-----DRVITITDSLARRELY- 409
Qy 743 IHTLMQETETDEISGNKLVTTTLRYARGAWDGRERFRGFGVVEQTD-----SHOLAQ 794
Db 410 -----TEGE--GGLKRVVKKEHADGSIITRSEYDEAG-RLKQATDAAGRTESYSLHMAS 459
Qy 795 GNAPERTPP-----ALTKWYATGLFVIDNALSTEYWRDQAFAGSPRFTT 841
Db 460 GAVTAVTGPDRTVRYGYNQRQVTSVTPDGL-----RSSREY----- 498
Qy 842 WQONKDVLPEDNDSRYNFRALKGQLLSELYGLDDSTNKHVPYTVTFPRSQVRLOH 901
Db 499 -----DEKGLAAETSRSETTR---YSYDDPASE-LPTGIQD-----A 533
Qy 902 TDSRYPVLWSSVVSERNYHYERIASDPQCSNIT-LSSDRFGOPKLSVQYPRRQPAI 960
Db 534 TGSTKQMAWS-----RYGQLITFTDCSGYTRYEYDRYG---QIAV-----HREEGI 578
Qy 961 NLYPDTLPDKLANSYDDQORQLRLTYQSSSWHLLTNNTVRLGLPDSRSDI--FTYG- 1017
Db 579 STYSSYNPRQLYSQDAQGRETR--YEYSAAGDLT-----AIVAPGSRSEIQYDAGWK 631
Qy 1018 AENVPAAGL-----NLELLSDKNSLIADDK--PREYLOQOQKTAITDQON----- 1059
Db 632 AVSTTQGLTRSMGYDAAGRIITVLTNENGSQSTFRYPDVRDLTEQRG--PDGRTQRYHD 689
Qy 1060 -TPTLQTPTRQALIAFTETTVFNQSTLSAFNGSIPSKLSTLLEAGYQQTNYL--FPRT 1116
Db 690 LTKGLTQSEDEGLVTLMHYDASDRITRHTVNGD-----PAEQWQYDHBGMLTTLTSH 741
Qy 1117 GEDKVVVAHRYGTYDG-----TAAQFWRPQ-----KQSNLTQLTGK 1151
Db 742 SEGHRVSVHYGYDDKGELTGERQTVENPETGEMLWEHETGHAYSEQGLATRQEPDGLPPV 801
Qy 1152 ITLWDANYC-----VVOTRD-----AAGLTTSAKYDWRFLTP 1185
Db 802 EWLTYSGYLAKMKLGGTPLVEYMRDLRHRETARSFGGEAYELATAMWTSQLASRHLNL 861
Qy 1186 VOLT---DINDNQHLLTLDALGRPITLRFWGTENGKMTGYSSPEKASPPSDVNAIEL 1242
Db 862 POLDRDYDNDNGQLIRISGPQSEYRY--SDTGLRTGVHT-----TAANLDIDI 910
Qy 1243 KKPLPVAQCQVAPESWMPVLSQKTFNRLAQDQWQKLYNARIITEDGRICTLAYRRWQS 1302
Db 911 -----PYATDPAGNRLPD----- 923
Qy 1303 QKAIPQLISLLNNGPRLPHSLTLTTRDYDHPDQIQIRQ---VVFSDGFGRLLOAAARH 1359
Db 924 -----PELHPDS-TLTA-----WPNRRIAEADAHVYRYDEVYGRLEAKTDRI 963
Qy 1360 EAGMARQENEGSLIINVQHTENRWATGRTYDKNKQPIRTYQPYPLNDWRYVNSDAR 1419
Db 964 PEGVIRMHDE-----RTHHYHDSQHLVFHTRIQ 993
Qy 1420 QEKEAYADTHVDPIGRIKIVITAG-WFRRTLFTPMFTVNEDENDT 1465
Db 994 HGEQVESRYLYDPLGR-----TGKRVRRERDLTGWMSLSRKPEET 1036
```

## RESULT 8

A99720  
RhsC core protein with extension [imported] - Escherichia coli (strain O157:H7, substrain  
C;Species: Escherichia coli  
C;Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 31-Dec-2004  
C;Accession: A99720  
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.  
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.  
DNA Res. 8, 11-22, 2001  
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and geno  
A;Reference number: A99629; MUID:21156231; PMID:11258796

A/Accession: A99720  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-1399 <HAY>  
 A/Cross-references: UNIPROT:Q8X9E8; UNIPARC:UPI00001653CB; GB:BA000007; PIDN:BA834152.1;  
 A/Experimental source: strain O157:H7, substrain RIMD 050952  
 C/Genetics:  
 A/Gene: EC60729

Query Match 2.1%; Score 164.5; DB 2; Length 1399;  
 Best Local Similarity 19.6%; Pred. No. 0.067;  
 Matches 305; Conservative 166; Mismatches 570; Indels 513; Gaps 80;

QY 18 GGALTG-----MGEALTPGPGMAALSPLPISAGRGVAPAFITLNNYSGAGNSPGLG 71  
 DB 41 GGTVSGHPVNPGLAKVLPGETD--IALPGLPILSRITYSSYRT---KTPAPVSLGPG 95  
 QY 72 WDCNV---MTIRRT-----HFGVPHYDETDTFLGPEGEVLVADQPRDESTL----- 116  
 DB 96 WKMPADIRLQRLDNTLILSDNGRSLYFE-HLPPGEDG-----YSRSESLVVRGVA 147  
 QY 117 ---QGILGATFTVTVGRSLSHPSRLVWQPKTKTDFWLIYSPDQVHLIG---KS 170  
 DB 148 KLDEGHELAALWQALPEELRSLPH-----RYLATNSPQGPWMLLGMCEV 192  
 QY 171 POARISNPSQTTQTAQWLLLEASVSSRGE-QIYYQYRAEDDTGCEADEITHLQATQRYL 229  
 DB 193 PEADEVLPAFLPPYR---VLTLGLVDRFGTQTFHRAAAGEFSG---EITGVTDGAGHFR 246  
 QY 230 HIVYGNRTASE-----TLPLGL-----DGSAPSQADMLFYLVFDYGER 267  
 DB 247 LVLTTQARAEARQQAISGTEPSAPFDLTPGYTEYGRNGIRLSAVMLTH-----DPEY 302  
 QY 268 SNNLKTTP---AFSTGSMWLCQDRFSR---YEYGFETRRLC-----RQVLMYHHL 314  
 DB 303 PENLPAAPLVRYGWTTPRGELAVVYDRSGKVRSTYDDKYGRMVAHRHTGRPEIRYR- 361  
 QY 315 QALDSKITEHNGPLVSRLLINLNDYSAITLVPVR-VGHEQ-DGNVVTLPPLLAYQD 372  
 DB 362 -DSGRVTEQLNPAGLS-YTYQYBKORITITDSLNRRVHLTQEGGLKRVVKE----- 414  
 QY 373 FSPRHHAHQFMDVLANFNALQIWMQLVKGELPGLLYQDKGAWMYRSQRLEIGSDA 432  
 DB 415 -----HADGSVTQSQFQDAVR-----LRAQ-----TDAAGRTTEYSPDV 448  
 QY 433 VTWEKQPLSVIPSLQNASLVDINGQDLWDITGP-GL-----RGVHSQRPDG 481  
 DB 449 VT-----GLITRITTPDGRASAFYNNHHSQLT-SATGPDGLFIRREYDEWGLRIQETAPDG 503  
 QY 482 SWTRFTPLNALPVREYTHPRAQL-----ADLMGAGLSDLVLIGPKSVRLVYANTR-DGPAK 536  
 DB 504 DITRY-----RYDNPHSDLPCLATEDATG---SRKMTWSRYGQLLSFTDCSGYTRY 552  
 QY 537 DVVQSGDITL-----VPFGADPR-KLVAFSDVLGSGQAHLEVSATKVTCPNLRG 587  
 DB 553 DHDRFGQMTAVHREGLSQYRAYDSRQLIAVKDTQGHETRYEYNAAGDLTT----- 604  
 QY 588 RFGQITLPGSPQATPENSQAQVYLAIDLDSGPTDLIYVHTNRLDIFLNKSGN--GFAEP 645  
 DB 605 -----VIAPDGSRTGQYDAGWKAICTTQGGLTSMYDAAGRILRTSENGSHTTFRYD 659  
 QY 646 VTLRFPGLRFDHTCQLQMDVQGLGVASLILSVPHMSPHWRCD-----LTN-----M 694  
 DB 660 VLDRLIQETGDTGRYHHDLTG-----KLIRSEDEGLVTHWHVDEADRLTHRTVKGETA 715  
 QY 695 KPWLLNEMNNMNVHHTLR-YRSSSQFWLDEKAAALATGQTPVCYLPFPFIHTLMQETED 753  
 DB 716 ERWQYDERGWLTDISHISEGRVTVHYGYDEKGR--LTGERQTVHHQTEALLWQHETRH 773  
 QY 754 EISGNKL-----VTLRLYARGAWDG-----RERFRFGVYVEQ 786  
 DB 774 AYNAQGLANRIPCPSLPAVEWLTYSGLWAGMLKGLDPLVDFTDRDLRHLRFRFGYEL 833

QY 787 TDSHQLAQGNAPERTTPALTKNWYATGLPVIDNALSTEYWRDQAPAFSPRFTTWQDNK 846  
 DB 834 TTAY-----TPAGQLQSHLNSL-----QYDRD-----YTWNNDN- 862  
 QY 847 DVPLTPEDDNSRYWFRNALKGQLLR-----SELYGLDDSS---TNKH-----VPYT 888  
 DB 863 -----GELIRISSPQRTSRYSYSDSGRLTGVTHTAANLDIRIPYA 902  
 QY 889 VTEFRSQVRRLQ-HTDSRYFVLMSSVVERNVHYRIASDPQCSQNTLSSDRFGQPLKQ 947  
 DB 903 TDPAGNRLPDPPELHPDSTLS-MWPNRNIARDAHY-----LYVRDHRGLTEK 948  
 QY 948 LSVQYPRRQOPAINLYPDLTPKLLANSYDDQORQLRLTYQSSWHHLTNNT----- 999  
 DB 949 -----TDLIPEGVIRT---DDETRH---YHYDSQRLVHYTFTQYEEPL 987  
 QY 1000 --VRVLGLPDSTRSDIFTYGAENVPAAGLNLELL-----SDKNLSIADDPREYLG 1048  
 DB 988 VESRYLYDPLGRRVAKRVRERDRLTGWMSLSRKPKQVWYGMWDGRLTTIQNDRTR--- 1043  
 QY 1049 QQKATYTDQONTTPTLPTFRQALIAETETTVFNQSTLSAFNGSIPSDKLSLTLEQAGYQ- 1107  
 DB 1044 -IQTIYQPGSFPLIRVETATGELAKTORR-----SLADALQQSGGED 1085  
 QY 1108 -----QTNLYFPRTGED-KVMVA----- 1124  
 DB 1086 GGSVFPFVLVQMLDRLESEILLADRVSESRRLWASCGLTVAQMSQMDPVYTPARKIHL 1145  
 QY 1125 ----HHG-----YTDYGTAAQFWRP-QKQSNLTQITG-----KITLIWD 1157  
 DB 1146 YHCDHRLPLALISKEGATEWCAEYDEWGNLLNEPNHQLQLIRLPQQYDEESGLYN 1205  
 QY 1158 AN-YCVVVO-----TRDAAGLTTSKADWRF-----LTPVQLTDINDNQHLLTDALGRPT 1208  
 DB 1206 RHRYDPLQGRYITQDPIGL-----KGMWYFYQYPLNPVQY-----IDSMG--LA 1248  
 QY 1209 LRFWGTENGWGTQYSS-PEKASPPSPSDV-NAAILKLPVQAQCVQYAPESMMPVLSOK 1266  
 DB 1249 SKYGLHNG---GYGARPNKPTPDSPKFPDIAKQRLRYPIDQA-----SSAPNVFKT 1299  
 QY 1267 TPNRLAQQWQKLYNARIITEDGRICITLAYRRVWQSKAIPQLISLLNNGRPLP 1320  
 DB 1300 FFRALSPYDY-----TLCKRWVK-----PNLTCTPQDDPQYP 1332

RESULT 9  
 S46217  
 protein-tyrosine-phosphatase (EC 3.1.3.48) type sigma precursor - rat  
 N/Alternate names: leukocyte common antigen-related phosphatase  
 C/Species: Rattus norvegicus (Norway rat)  
 C/Date: 07-May-1995 #sequence revision 03-Nov-1995 #text\_change 09-Jul-2004  
 C/Accession: S46217; S51174; A49104  
 R/Zhang, W.R.; Hashimoto, N.; Ahmad, F.; Ding, W.; Goldstein, B.J.  
 Biochem. J. 302, 39-47, 1994  
 A/Title: Molecular cloning and expression of a unique receptor-like protein-tyrosine-phosphatase  
 A/Reference number: S46216; MUID:94347119; PMID:8068021  
 A/Accession: S46217  
 A/Status: nucleic acid sequence not shown  
 A/Molecule type: mRNA  
 A/Residues: 1-1863 <ZHA>  
 A/Cross-references: UNIPROT:Q64605; UNIPARC:UPI000017705B; EMBL:L11587  
 R/Goldstein, B.J.  
 submitted to the EMBL Data Library, February 1993  
 A/Reference number: S51174  
 A/Accession: S51174  
 A/Molecule type: mRNA  
 A/Residues: 1-1788,'G',1790-1863 <GOL>  
 A/Cross-references: UNIPARC:UPI000008B4D4; EMBL:L11587; NID:q205134; PIDN:AA37656.1; PII  
 R/Yan, H.; Grossman, A.; Wang, H.; D'Eustachio, P.; Mossle, K.; Musacchio, J.M.; Silvenn  
 J. Biol. Chem. 268, 24880-24886, 1993  
 A/Title: A novel receptor tyrosine phosphatase-sigma that is highly expressed in the ner  
 A/Reference number: A49104; MUID:94043351; PMID:8227050  
 A/Accession: A49104

A>Status: preliminary; not compared with conceptual translation  
A:Molecule type: nucleic acid  
A:Residues: 1-596, 'R', 598-603, 'I', 967-1788, 'G', 1790-1863 <YAN>  
A:Cross-references: UNIPARC:UPI000005E5D5  
A:Experimental source: brain  
C:Note: sequence extracted from NCBI backbone (NCBIP:139669)  
C:Superfamily: leukocyte antigen-related protein; fibronectin type III repeat homology; ogy  
C:Keywords: alternative splicing; glycoprotein; phosphoprotein; phosphoric monoester hyd  
F:1-26/Domain: signal sequence #status predicted <SIG>  
F:27-1863/Product: protein-tyrosine-phosphatase #status predicted <MAT>  
F:149-209/Domain: immunoglobulin homology <IMM1>  
F:246-300/Domain: immunoglobulin homology <IMM2>  
F:318-400/Domain: fibronectin type III repeat homology <FN3A>  
F:413-499/Domain: fibronectin type III repeat homology <FN3B>  
F:511-592/Domain: fibronectin type III repeat homology <FN3C>  
F:1244-1863/Domain: leukocyte common antigen cytosolic domain homology <LAC>  
F:1331-1552/Domain: protein-tyrosine-phosphatase homology <PTP1>  
F:1504/Active site: Cys (phosphocysteine intermediate) #status predicted  
F:1510/Binding site: substrate phosphate (Arg) #status predicted  
F:1795/Active site: Cys (phosphocysteine intermediate) #status predicted  
F:1801/Binding site: substrate phosphate (Arg) #status predicted

Query Match 2.0%; Score 161; DB 2; Length 1863;  
Best Local Similarity 19.1%; Pred. No. 0.18;  
Matches 336; Conservative 188; Mismatches 613; Indels 620; Gaps 87;

Qy	9	ITELSLPQGGG---AATGMBALPTGPDGMAALSPLPISAGRYAPAFATLNTNSGAGN	65
Db	313	ITVKSLPKAGPTPVVTENTATSIITVDSGN-----PDVVS-----YYVIEYKSKQD	360
Qy	66	SPFGLGDCNVMVITRRTHFGVPHYDETTTL-----GPEGVLV-----ADQP	110
Db	361	GPYQIKED---ITTRYSIGGLSPNSEYEIWMVSAVNSIGQPPSESVTRTGEQAPASAP	417
Qy	111	RDESTLOGINCATFTVGYRSRLESFHSRLEYWPKTKTDFWLIYSPDGQVHLLGS	170
Db	418	RN---VQARMLSATTMIVQWEEPEPN-----GLIRGYRVYTYMEPEHPVGNW	462
Qy	171	PQARSINSPQTTQTAQWLEASVSRGQIYYOYRAEDDTGCEADEITHLQATAQRYL-	229
Db	463	QKHNVDLSLIT-----VGLLEDEYTYVTVLAFTSVGDGLSDPIQVKTQQQVP	512
Qy	230	-HIVYGNRTASSETLPGLDGSAPSQADWLFI-LVPDYGERNNLKTTPAPFSTTGSWLCRQ	287
Db	513	GQPMNLRABAKSETSIGLSWSAPRQESVIKVELLPREGDRGRCR---TFDPTTAFVV-E	569
Qy	288	DRFSRYEGFEIRTR-----LCQVLMYHHLQA-----LDSKITEHNGPTLVS	331
Db	570	DLKPNTYAFRLAARSPQGLGTAFTAVVCQRT-----LOAKPSAPPQDVKCTSLRSTAI--	622
Qy	332	RLIL-----NYDESAIA-----STLVFVRVGVHEDQGNVVTLPPLLEYQDPS	374
Db	623	-LILLEALEKTEYRTAVAYTEVGPGEPSVVVR-----TDEDVPSAPPKRVAEALN	676
Qy	375	PRH-HAHQPMQMDVLANFNAIQRWQLVDLKGEGPLGLLYQDKGAWYRSAQRIEIG-SDA	432
Db	677	ATAIRVLRSPPTGRHQGIRGYQVHYVRMEGTEA-----RGPPRIKIDIMLADA	725
Qy	433	VTEKQQLSVIPSLQS-----NASLVDINDGQ---QLDWITGPGRLGRTHS-----ORP	479
Db	726	---QEM---VITNLQPTAYSIITVAAYTMKGDGARSFKVVTWTKGAVLGRPTLSVQQT	778
Qy	480	DGS-----WTRFTPLNALPVEYTHPRAQLADLMGAGLSDLVL-----IGPKSVRLY	525
Db	779	EGSLLARW-----EPADAAE-----DPVLGYRLQFGREDAAPATLELA	817
Qy	526	ANTRDGFA---KGKDW-----QSGDITLPVPGADPKRLVAFSDVLGSGQAH	570
Db	818	AWERFPAAPAHKATVTVFLAARGRAGLGEASALSIPEDAPR---GPPQILGPAG---871	
Qy	571	VEVSATKVTCT-W-----PNLGRG---RFGQPIITLPFGSQPATE-----FNPAQVYLADLD	616

Db	872	NVSAGSVILRWLPVPVPAEGNGAIIKVTVSVRAGTGGPATETELAAAAQGAETALTQ	930
Qy	617	GSQFTDL-----IYVTHNRLLDIFLNKSGNG-FABEPVTLR-----	649
Db	931	GLRPTAYELRVRAHTR-----GPGFSPPLRVRLARDVPSPKPKVKIMKTSVL	982
Qy	650	-----PPEGLRPHDTCQLQ-----MADVQGLGVASLIISVPHMSPHHWRCDLTNNKPKWLLNE	701
Db	983	LSWEPFDNYSPTPKYKIQNGLYGLDVGRTTKLLI---THLXPH-----TFYNFVLNTR	1033
Qy	702	MNNMGVHHHTLRVSRSSQFWLDSKAAA-----LTTGTQTPV-----CYLP--	740
Db	1034	GSSLGGLQQVTAARTAFNMLSGKPSVAPKPDNDGSIIVYLPDQGSPTVQNYFIWVPLR	1093
Qy	741	-----PPIHTLMQTEDEISGNKLV-----TTLRYARGA-----	770
Db	1094	KSRGQGFPI---LLGSPEDMDLEELIQLDLSRLQRSLRHSRQLEVPVPYIAARFSILPA	1149
Qy	771	--WDGRERPRGFGYVEQTDHQLAQN-----APERTPALTKWYATGL-----	814
Db	1150	VFHGPNQKQYGGF-----DNRGLEPGRHYVFLVLAQLQKNEPTFAASPFSDPQLDNDP	1203
Qy	815	--PVIDN-----ALSTEYWRDDQAFAGFSRFTTWOONKDV-P	849
Db	1204	POFIVDGEGLIWIWGPVLAIVFIIICIVIAILLKPKDSEKRSKDSERPTKCLANNADLP	1263
Qy	850	LTPEDDNRYSFNRALKGQLLRSELGYLDDSTNNKHVYTVTFRSQVRRLOHTDSRYPL	909
Db	1264	HHPKDPPVEMERINFQTPGML-----SHRPIPTDMAEHMERLKANDSL---	1306
Qy	910	WSSVVSERNHYRIASDDPOCSQNTILSSDRFQPLKQLSVQYPRRQOQAINLYPDTL--	967
Db	1307	-----KLSQYESI--DP-----GOQFTWEHSNLEANKPKNRYANVAYDHSRVI	1349
Qy	968	-----PKLLANSYDDQQLRLTYQSSWHHLTNNTVRLVGLPDSRSDIPTGYA	1018
Db	1350	LQPLEGIMGSDYINANTVGYRRQ-----NAYIATQGLPETFGDFWRWVW	1395
Qy	1019	ENVPAGGLNELLSDNKSLIADDPKPREYLGQOKTAYTDGQNTTLPQTPTTQALIAFTETT	1078
Db	1396	EORSATVVMTRLEESRVCQDQ-----YWPNRG-----TETY	1428
Qy	1079	VFNQSTLSAPNGSIPSKLSTTLTLEAGYQQTNYLFPRTG-EDKVVVAH-----HGYT	1129
Db	1429	GFIQVT-----LLDTMELATFCVTFESLHKGSSSEKREVRHFQFTAWPDHGVP	1476
Qy	1130	DYGTA-AQFWRPQKQNTQLTKITLIWDANY---CVVV-----QTRDAAGLT	1173
Db	1477	EYTPFLAFRRVKTCTMPPDAGPVVHCSAGVGRGTGCFIVIDAMLERIRTEKTVDDVYGHV	1536
Qy	1174	TSAKYDWRFLTPVQLTDINDNQHLITLDALGRPITLRFWGTENGKMTGYSSEKASFPSP	1233
Db	1537	TLMSQRNYM--VQTEQOYSFIHEALLEAVG-----CGNTEVPARSLEYIYQKLAQVEP	1588
Qy	1234	SDVNAALTELKPLPVAOCQVYAPESMMPVLQKTFNRLASQDWOKLXNARIIT-----	1286
Db	1589	GEHVTGMELE-----FKELAS-----SKAHTSRFITASLPCNK	1621
Qy	1287	-----EDGRICITLAYRWVQSQKAIPLQILSLNNGPRPLPHSLITITTDYDHP	1335
Db	1622	FKNRLVNILPYESSRVC-LQPIRGVSGSDYI-----	1651
Qy	1336	EQIIRQOVFSDGFRLLQAAARHEAGMARQNRDGLIINVOHTENRWAVTGYEYDNK	1395
Db	1652	-----NASFIDGY-----ROQKAYIATQ---GFL---AETTEDFW---RALWENN	1687
Qy	1396	GQPIRTYQPYFLDNWRVYVNSDSARQ-----EKEAYADTHVYDPIG-----REIKVIT	1442
Db	1688	STIV-----VNLTKLRMGREKCHQYWPASARSARYQYFVVDPMAYNNPQVILREPKVTD	1742
Qy	1443	AKGWFRTIL-----FTPW	1455
Db	1743	ARDQGSRTVRQFQFTDW	1759





Qy 1421 EKEAYADTH-VYDPI-GREI 1438  
Db 1210 ESGLYNRRHYDPLQGRYI 1229

RESULT 11  
AH2493  
hypothetical protein all7128 [imported] - Nostoc sp. (strain PCC 7120) plasmid pCC7120a1  
C:Species: Nostoc sp. PCC 7120  
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
C:Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 09-Jul-2004  
C:Accession: AH2493  
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, N.; Shimizu, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.  
DNA Res. 8, 205-213, 2001  
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena PCC 7120  
A:Reference number: AB1807; MUID:21595285; PMID:11759840  
A:Accession: AH2493  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-3083 <KUR>  
A:Cross-references: UNIPROT:Q8YL10; UNIPARC:UPI000000CEEDB; GB:BA000020; PIDN:BAB78212.1;  
A:Experimental source: strain PCC 7120  
C:Genetics:  
A:Gene: all7128  
A:Genome: plasmid

Query Match 2.0%; Score 158; DB 2; Length 3083;  
Best Local Similarity 18.8%; Pred. No. 0.66;  
Matches 276; Conservative 163; Mismatches 507; Indels 520; Gaps 73;

Qy 1 MONSDFSIITSLPKGGGATKMGREAL-TPTGPDG-----MAALSPLPIISAGRYA 52  
Db 725 IEAAAEVNVAASAGAGGGIIGTGLNKKDPTPGDKVRGNEFVQLLNPNPIEMFEDASGLV 784

Qy 53 PAPTLYNSGAGN-----SPFG-----LQWDCNVWIR----- 80  
Db 785 QAYLMAYAKVAGKVVKRIESPVKLLPGYKVSFETPPQLHLATDIGGGLNLRNMPNAAA 844

Qy 81 -----RRTHFGV-----PHYBTD 94  
Db 845 REIINTEGAEVFTTDDGKLTYSAFNIPTQYSGVSKIIADGGTYNDTIEIKDIEISA 904

Qy 95 TPLGPEGLVAVDQPRDESTLOGINLGATFTVTGYRSRLESFHSRLEYWQPKTKGTDF 154  
Db 905 DLKGAGEDLYCGGSD--TIRG--GAD-----WDRLYGGDRDD 940

Qy 155 WLIYSPDQVHLGKSPQARISNPSQTTQTAQWLLASVSRGEQIYYQYRAEDDTGCEA 214  
Db 941 P-VYGGDDDWLDG-GAGADILNGAGPDT-----ASYTSATSAISINLVQVSTGDAA 992

Qy 215 DEITHLL-QATAQRYLHTV-----YGNRTASSETLPG-----LDG--- 248  
Db 993 DDVQSQEQIVGSRYYDTLIGDEDNNEFDGEGNDFTSGGAGDRLSPGWDVDDGTG 1052

Qy 249 -----SAPSOA-----DMLFYLVDYGERSNLKTTPAPSTGTGWLCKRD 288  
Db 1053 TDTLVIDYSSLTQAVAMSELDPNTSDWFFVAVANAYGIGA-PIKTD--INVSGN----- 1103

Qy 289 RFSRYEYGFETRRLCRQVLMYHQLQALDSKITEHNGPTLVSLRLILNYDESATASTLVF 348  
Db 1104 -----YHATLSAD-----GLTVAGSGILGSGNS--GNQGLV 1132

Qy 349 VRRVHEODGNVVTLPPELAYQDPSPRHAAH--WQPMDV--LANFNAIQRWQLVDLK-- 402  
Db 1133 VKXI-HSSDPAVRIPNNQV-YQPLLEDGSKVWVSGDSSTWIANTNGTQVRQTKLSIN 1190

Qy 403 -----GEGPLGLLYQDKGAWWYRSAQRAGEI-----GSDAVTWKMKQPLSPISLQSN 450  
Db 1191 IGYGDGDLATISEDGSGTIAMLSKRNNDKFTYTFIFIANADGKLNQIN-----IPTSGG 1246

Qy 451 ASLVINDINGDG-QLDWVITGPGRLGRVHSQRPGSWTRFTPLNALPVETHPRRAQLADLMGA 509

Db 1247 VRELDLSADGSKITWSQDG---GY---GPGGVWVANT-----DGTNIRELSGNLYGY 1292  
Qy 510 GLSDLVL-----IGPKSVRLYANTRDG--FAKGKVVQSGDITLTPVPGADPR 554  
Db 1293 NINPSISADGSTVVMAGYQAGYASTNLXAATDGSRFVVVPTNTEEVGEFAQAQSLAGDSR 1352

Qy 555 KLV-----AFSDVLGSGQHILVFSATKVTCPNLRGRFQOPITLPGFSQPAT 603  
Db 1353 RVVPTKFGSDYSLYVGDIDGIEPQILIDASS-----PNIGIRGRHALSSYVLDGVRYN 1406

Qy 604 EFNPA-----QVYLADLDSGSGPTDLYVHTNRLDIF-----LNKS--GNGFAEP 645  
Db 1407 SFDPATSGSEIYT-----WGPSRIYRNSFERFDIIGTRYGDELPGNLDLMSLGGGGAD- 1460

Qy 646 VTRFPPEGLAFDHTCQLQMAADVQGLGVASLILSVPHMSPHHWRCDLNTMKPFLNENWNN 705  
Db 1461 -TLK--AGLGDD---IYILDTONAGGSQIE-----DAGTD--TLRLITRN 1498

Qy 706 MGVHHTLYRSSSQFWLDEKAAALTTCTQTPVCYLPFIHTLMQPTET-----EDEISGN 758  
Db 1499 PGATNTPRITDA-----DLSLAVPTTG-----IFGMRRAGTSLIIDLNKDGIAAS 1543

Qy 759 KLV-VTTLRYARGAWDGRERFRGFGYVEQTDHQLAQGNAPERTPPALTKNWTATGLPVI 817  
Db 1544 KTDLTILNFPDVTGTG-----AGTGFITETVANLAGAE-----ILSK-----LQVG 1583

Qy 818 DNALS-----TEYWRDDQAFAGFSRFTTW--QDNKDVPLTPDDNSRY----- 859  
Db 1584 DDTISGSAADFDIGWLSNNDTLSCGAGNDILWQDGNDF-LNGEDGNDLSQGGNGNDTLT 1642

Qy 860 --WPNRAKQLRLSELYGLDDSTNKHVYVTVTFRSQVRLQHTSDSRYPVLMSVSVESR 917  
Db 1643 PGWGNVDVDDG-AGTDVLVD-----YSNLNTRA-----VANRTLSGTS 1680

Qy 918 NYHYERTASDPOCSQNTLSSDRFGQBLKQLSVQYPRQOPAINLYDPTLPDKLLANSYD 977  
Db 1681 GNYLQKF-----FIGNAYGLGTPLKIRET-----NSVSDKFALSA-- 1715

Qy 978 DQORQLRLTYQBSWHLTNNTVRVLGLPDSRSDIFTGYAENVPAAGLNLLELSDKNSL 1037  
Db 1716 -----DGTTVAYTYINNDPANGWLKIKDDSGGL 1746

Qy 1038 IADDKPREYIG---QOKTAYTQDQNTTPTQTPROALIAFTETTVFNQSTLSAFNGSIP 1093  
Db 1747 VKIDEIATEIALSDTGEKIAWSG---RVVAVANTNGTEKIRINLANNINGYI 1796

Qy 1094 SDKLS-----TTLBQAGYQQTNYLFPRTGEDKVVVAHHGYTDYG 1132  
Db 1797 SLSLSGDGSQVSWNNGNOLLVANTDGTNIREITQSSSTKSLSENGSQIHWAGYQG-EKYG 1855

Qy 1133 -----TAAQFWRPQKQSNLTQTKG--ITLIWDANYCVVVQTRDAAGLTTSAK-YDWRP 1182  
Db 1856 IWSNSTSTSLPVVKSVDGNLSLSSSDGKAIKWDRYFLSVSSTNSTEIQOVAESYDFRV 1915

Qy 1183 L-----TPVQLTD-----IN-DNQ 1195  
Db 1916 VGSSEPVLAADGAKVAFIKAINADNQ 1941

## RESULT 12

AF1489  
cell wall-associated protein precursor wapa (B. subtilis) homolog lin0454 [imported] - L  
C:Species: Listeria innocua  
C:Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 09-Jul-2004  
C:Accession: AF1489  
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker, D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussauguet, O.; Entian, K.D.; Fsihi, H.; Jones, L.M.; Karst, U.  
Science 294, 849-852, 2001  
A:Authors: Kraft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma, O.K.; Schluster, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland, A.; Title: Comparative genomics of Listeria species.

A;Reference number: AB1077; MUID:21537279; PMID:11679669		
A;Accession: AF1489		
A;Status: preliminary		
A;Molecule type: DNA		
A;Residues: 1-2167 <GLA>		
A;Cross-references: UNIPROT:Q92EK5; UNIPARC:UPI00000CC234; GB:AL592022; PIDN:CAC95686.1		
A;Experimental source: strain Cl1p1262		
C;Genetics:		
A;Gene: lin0454		
Query Match 2.0%; Score 156; DB 2; Length 2167;		
Best Local Similarity 18.9%; Pred. No. 0.49;		
Matches 244; Conservative 163; Mismatches 399; Indels 484; Gaps 69;		
QY	314	LQALDSKLTENHPTLVSRLLINVDSEAIASLTVFVRVCGEODGNVTVLPPIELAYQDF 373
Db	416	VQWASGSRVNGFRGLTNKINWKVASE-----NNKNPYLELVNYT-- 460
QY	374	SPRHHAHQPMQDVLANFNARWQ-LVDLKGEGLPG-----LLYQDKGAMWYRSARLGE 427
Db	461	-----YAQPEKPTVKTSSNGVGTGYMDLSKAVPGATSNIVISNGYKYEYNTK----- 512
QY	428	IGSDAVTW-----EKMOPLSVIPSLQSNASLVDINGQQLDWITGPGLRGYHSQRPDGSW 483
Db	513	--STATTWSTGKKIKFTN-----DEIAKGEFEF-----HH-----DGKG 545
QY	484	TRFTPLNALPVEYTHPRAQLADLMGAGLSDLVLGPKSV-----RLYANTRDGPAKGKDV 538
Db	546	TDEFA-----LDPRAYENAFQAGST-----FGLNLTRYLFRVQAVFPFGESPNSDL 592
QY	539	V-----QSGDITL-----PVPGADPRKLVAFSVDVLGSGQAHLVE 572
Db	593	VPAYMPIEKPOAPVAKAYSNAHKETGYVELNWEKSPMADGYKVLVE-----NGKAY-EE 646
QY	573	VSATKVTCPNLKGRPGQPTILPGFSQATERNPAQVYLADLDGSG---PTDLIYVHTN 629
Db	647	YDVGAETKMTQNGI-----PTKEBIAEGKYALHHDGKAELAKDPSPVYTN 695
QY	630	-----RLD-----IFLNKSGNGP-----APVTLRPEGLRFDHTCOLQADVQGLGVA 673
Db	696	SGGNYKERTWYFRVIAAYQAGNNVTSVQSEPAFSPSEALN-----KQLGMV 743
QY	674	SLILSPHMSPHWRCDLTNMKPWLNNMNNMVHHTLYRSSSQFWLDEKAAALTTGQ 733
Db	744	DYWTSP-----VRGEVNAWNGFLFHEF-----DNLSGRGPSINVR 783
QY	734	TPVCYLPFPPIHLWQTEDEISG-----NKLVTLLYARGAWDGRERFRFGF 782
Db	784	T-----FNSQDDATGIPGKGWTSITLLEKLV-----EEENGNIIV 816
QY	783	YVEQTD8-HQLAQGNAPERTPPA-----LTKNWYATGLPVIDNALSTEWYRDDQAPAGFSP 837
Db	817	WVESDKKHFRFTKGDYKYEAPPGLYSBITKN-----ADGYLKIEEDKSETRFL-VDG 867
QY	838	RFTTWQDNK-----DVPLTPEDDNSRYWFRNALGQLLRSELYGLDDSTNKHVPYT 888
Db	868	RLKSEKDTGNELTYETDGLKSLRDASGRTVTLTYEGELVK-ELVGPE-----RKISYT 923
QY	889	VTE-----FRSQVRRLQHTSDRSPVILWSSVVERSN-----YHYERIASDPQC 930
Db	924	YNDQELISSSTARGKLYRYGYTDG-----LLTSYDPKPKTEKPYTTFAYE----- 971
QY	931	SQNTILSSDRFGPLKQLSVQYPRRQPAINLPDPLDKLLANSYDDQOQLRLTYQOS 990
Db	972	EEKLTEITDPVG-----KKTLSYDKASQOQTLTNEKK 1004
QY	991	SWRHLTNNTVRVLGLPDSRSDIFTYG-----AENVPAAGLNL-----ELLSDKNSLIADDK 1042
Db	1005	-----KKTIVSYNDAGNPKKEIVDADGLKLTITYTESNNLVKEVN 1045
QY	1043	PREVLQOQKT-AYTDGQNTTLPQPTFOALIAF-TETTVFN-----QSTLS 1086
Db	1046	PK---QOETVAYDADGNIT-----KATDAYGTESYTYNNNDVNTSSTDTEGRKTTVT 1095
RESULT 13		
E90886		
RheE core protein with extension [imported] - Escherichia coli (strain O157:H7, substrain C;Species: Escherichia coli		
C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 09-Jul-2004		
C;Accession: E90886		
A;Reference number: A99629; MUID:21156231; PMID:11258796		
A;Experimental source: strain O157:H7, substrain RIMD 0509952		
C;Genetics:		
A;Gene: ECe2061		
C;Superfamily: rhesF protein		
Query Match 2.0%; Score 155.5; DB 2; Length 1400;		
Best Local Similarity 19.6%; Pred. No. 0.25;		
Matches 266; Conservative 139; Mismatches 451; Indels 498; Gaps 71;		
QY	406	LPGLLYQDKGAMWYRSARLGEIGSDAVTWKMQPLSVIPSLQSNASLVDINGDGLDW 465
Db	70	LPFILSRTSYSTRTPAPVGIKPG---WK--APSDIRLQLRDDALVLDNDGGRSHFE 124
QY	466	ITGPG-----LRGYHSQRPDGSTRPTPLNALPVEYTHPRAQLADLMGAGLS 513
Db	125	PLLPGEAVYSRSESMVLVRGKAAQPDGH-----TLARLWGLALPPD 165
QY	514	LVILGPKSVRLYANTRDGPAK-----KDVVQSGDITLTPVGADPRKLVAFSVDVLGSG 566
Db	166	IRL-----SPHYLATNS--AQGPWILGMSRVERPGAEDVLPALPPYRVLTGMADRFRT 219
QY	567	QAHLEVESA-----TKVTCMPNLGRGR----- 588
Db	220	LTTRREAAGDLAGEITGVTD----GAGREFRLVLTVTAQBAEARKQHTASLSPDTPRP 275
QY	589	---FGQPIITLPGFSQPAFENPAQ-----VYLADLDGSGFTDLIYVHTNRLD 632

Db 276 LSDAPPDTLP-----TEYGPDRGIRLSAVWLTHDPAYPSLPQAPLARYTYTEAGELL 330

Qy 633 IFLNKS-----GNGFAPSVTLRFPPEGLRFDHTCQLQMAVQGLGVAS 674

Db 331 AVYDRSNTQVRAFTYDAQHPGRMVAHRVAGRPENRY-----RYDDTGRV-----VEQLNPAG 382

Qy 675 LILSVPHMSPHHWRCDLTNMKFWLLNENNNMGVHHTLR--YRSSQFWLDEKAAALTG 732

Db 383 LSVRYQVEQDRITVTDSLRREVLTHT--GGAGLKRVVVKSLADGVSHTSGYDAAGRUTA 440

Qy 733 QTPVCYLPFPPIHTLWQTEDEISGNKLVTLRVARG-----AWDGREREF----- 778

Db 441 QT-----DAAGRTVEYGLNVVSGDITDITPDGRETFYNDGNQL 481

Qy 779 -----RFG-----YVEQT-----DSHQLAQGNAPERTPPATKN-----W 809

Db 482 TAVVSPDGLESRRAYDEPGLRVSETSRGCDVIRAYDN-PHSELPAITTDATGSTRQMTW 540

Qy 810 --YATGLPVID-NALST--EYWRDDQAF-----GFSRPTTWQDNKDVLTPED--DN 856

Db 541 SRYGQLLAFTDCGYQTRYEYDRFGQMTAVHREEGIS--RYRRY-DNRGLTSVKDAQGHE 598

Qy 857 SRYWFNRA-----LKGOLLRSLEYGL-----DDS 880

Db 599 TRYEYNAAGDLTAVITPDGRSETQYDAGKAVSTTQGLTRSMYDLAGRITLTHENG 658

Qy 881 TNKHVPYTVTFERSQVRLQHTDSRYVPLVSSVSVESRNYHYERIASDPQCSON--ITL-- 936

Db 659 SRSEFTYDALDRLVQORGF--DGR-----TQRYHYDLTGKLTQSEDEGLVTLWH 705

Qy 937 --SSDRP-----CQPLKQLSVQVPRR-----QPAINLYPDTLPDKLLANSY 976

Db 706 YDESRLTRHTVNGEPAEQW--QYDEHGWLTEIHLSEGHQVAVHY-----GY 751

Qy 977 DDQORQL--RLTYQOSS-----WHLTNNTVRLVGL-----PDS-TRSDIFTYGAENVPA 1023

Db 752 DDKGLAGERQTVNPEGTCELLWQHETAHYNEQGLANRVTPDSLPVWEVLTYS----- 806

Qy 1024 GGLNLELLSDKNSLIADDPREYLGQOQTAYTDGQNTPLPTQALIAFTETTVFNQS 1083

Db 807 -----GYLAGMKL-----GGTFLVEFTDRL-----HRE 830

Qy 1084 TLSAFNGSIPSDKSLSTLEQAGYQNTYLPRTGEDKVVAAHGYDYGTAQFWRPQK- 1142

Db 831 TVRSFGNN--AYELTSTYPAHLQSQRL-----NSQVYDRDYDNNDGDLVRSQPRQT 883

Qy 1143 -QSNTQLTGKI--TLIWDANYCVVQTRDAAG-----LITSA 1176

Db 884 WEYGSATGRLESVRTLASDLDIRIPYAT-DPAGNRLPDPPELHPDSTLTAWPDNRIABDA 942

Qy 1177 KYDRFLTPVQLTDIN-----DNQHLITL----- 1200

Db 943 HYVVRHDEYGRLETKTDRI PAGVIRTDDBERTHYHYDSQHLRVFTRIQHGEPLVESRYL 1002

Qy 1201 -DALGRPTLFPWGTENGKMTGYS-----PKASFSPPSDVNAIELKPLPVAQCO-VY 1254

Db 1003 YDPLGRMAKVRWRER-DLTGWSLSRKPEVTWYGMGDRLLTVQ-----TDTTRIQTVY 1057

Qy 1255 APESWMPVLSOKTNRLAEOQWOKLYNARIITEGRICTLAYRRVWQSOKAIPQLISLIN 1314

Db 1058 EPGSFPLPIRVETENGEREKA-QRRSLAETLQOEGS-----E 1093

Qy 1315 NGPRLP-PHSLTLTTDRYDHPDQOIQOQVFSFGFGRLLQAAARHEAGMARQNRD--G 1371

Db 1094 NGHGVVPAELVRLDL-----EERADRVSSESRRAWLAQGLTVEQ-LARQVEPEYTP 1148

Qy 1372 SLIINVOHTNR-----WAVTGRTEYDNKGQPIRTYQPYFLND-WRYVSND 1416

Db 1149 ARKVHFYCHDRGLPLALISBDGNTAW-----RGEYDEWGNQLEENPYLHQPYRLPQO 1204

Qy 1417 SARQEKAYATHVYDPI-GREI---KVITAKGW 1446

Db 1205 HDEESGLYNNRYDPLQGRYITQDPITGLAGGW 1238

## RESULT 14

T18995

hypothetical protein C06B8.7 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004

C:Accession: T18995

R:Steward, C.

submitted to the EMBL Data Library, November 1996

A:Reference number: 219057

A:Accession: T18995

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-3036 &lt;WIL&gt;

A:Cross-references: UNIPROT:O17575; UNIPARC:UPI00001B740; EMBL:Z81463; PIDN:CAB03852.1;

A:Experimental source: clone C06B8

C:Genetics:

A:Gene: CESP:C06B8.7

A:Map position: 5

A:Introns: 102/3; 157/2; 288/2; 439/3; 542/1; 591/3; 798/3; 835/3; 1139/1; 1204/3; 1325/3

Query Match 1.9%; Score 153; DB 2; Length 3036;

Best Local Similarity 18.7%; Pred. No 1.3;

Matches 241; Conservative 164; Mismatches 459; Indels 428; Gaps 65;



Search completed: December 16, 2005, 14:43:37  
Job time : 64 secs



***This Page Blank (uspto)***

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

QM protein - protein search, using sw model

Run on: December 16, 2005, 14:37:07 ; Search time 13 Seconds  
(without alignments)  
763.846 Million cell updates/sec

Title: US-10-706-424-10  
Perfect score: 7901  
Sequence: 1 MQNSQPSITELSLPKGGGA.....WFTWNEEDNTAAEVKKVKM 1474

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 51470 seqs, 6736768 residues

Total number of hits satisfying chosen parameters: 51470

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications\_AA\_New\*

1: /cgm2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*  
2: /cgm2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*  
3: /cgm2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*  
4: /cgm2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*  
5: /cgm2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*  
6: /cgm2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*  
7: /cgm2\_6/ptodata/1/pubpaa/US11\_NEW\_PUB.pep.\*  
8: /cgm2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6031.5	76.3	1476	6	US-10-647-956A-4
2	139	1.8	2515	7	US-11-113-424-53
3	136	1.7	2769	7	US-11-113-424-14
4	119	1.5	1227	6	US-10-793-626-96
5	116	1.5	1348	6	US-10-995-561-624
6	116	1.5	2715	7	US-11-113-424-51
7	115.5	1.5	1192	6	US-10-858-730-72
8	114.5	1.4	1107	6	US-10-485-517-145
9	114.5	1.4	4419	6	US-10-821-234-1155
10	114	1.4	805	6	US-10-485-517-198
11	113.5	1.4	1206	6	US-10-858-730-73
12	113.5	1.4	4128	7	US-11-770-726-77
13	113	1.4	1907	7	US-11-000-463-250
14	112.5	1.4	915	6	US-10-647-956A-6
15	112.5	1.4	1432	6	US-10-510-386-218
16	112	1.4	2516	6	US-10-647-956A-2
17	110	1.4	2296	6	US-10-995-561-633
18	110	1.4	2355	6	US-10-995-561-623
19	110	1.4	2355	6	US-10-995-561-627
20	110	1.4	2384	6	US-10-821-234-1545
21	110	1.4	2386	6	US-10-995-561-626
22	108.5	1.4	782	6	US-10-972-053-2
23	108.5	1.4	782	6	US-10-972-053-8
24	108.5	1.4	1897	6	US-10-821-234-1635
25	107.5	1.4	1315	6	US-10-995-561-630

## ALIGNMENTS

## RESULT 1

US-10-647-956A-4  
; Sequence 4, Application US/10647956A  
; Publication No. US20050251878A1  
; GENERAL INFORMATION: Constant, Richard

; APPLICANT: Bowen, David  
; APPLICANT: Rocheleau, Thomas  
; APPLICANT: Waterfield, Nicholas

; TITLE OF INVENTION: DNA SEQUENCES FROM PHOTORHABDUS LUMINESCENS  
; FILE REFERENCE: 61645  
; CURRENT APPLICATION NUMBER: US/10/647,956A  
; CURRENT FILING DATE: 2003-08-26

; PRIOR APPLICATION NUMBER: CURRENT APPLICATION NUMBER: US/09/817,514

; PRIOR FILING DATE: CURRENT FILING DATE: 2000-03-26

; PRIOR APPLICATION NUMBER: US 60/191806

; PRIOR FILING DATE: 2000-03-24

; NUMBER OF SEQ ID NOS: 8

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 4

; LENGTH: 1476

; TYPE: PRT

; ORGANISM: Photorhabdus luminescens

; US-10-647-956A-4

Query Match

Best Local Similarity

Matches 1120; Conservative 122; Mismatches 224; Indels 9; Gaps 3;

Score 6031.5; DB 6; Length 1476;

Pred. No. 0;

1 MQNSQPSITELSLPKGGGAITGGEALTPGDPGMAALSPLPISAGRYAPFTLNYN 60

1 MQNSQPSITELSLPKGGGAITGGEALTPGDPGMAALSPLPISAGRYAPFTLNYN 60

61 SGAGNSPFGCGWDCNVMTIRRHFGVPHYDETDTFLGPEGEVLVA-----DQPRDES 114

61 SGAGNSPFGCGWDCNVMTIRRHFGVPHYDETDTFLGPEGEVLVA-----DQPRDES 120

115 TLQINLGATFTVTGVRSLRSHSFSLRYWQPKTTGTDFWLIYSPDQVHLLGKSPQAR 174

121 SLQINLGATFTVTGVRSLRSHSFSLRYWQPKTTGTDFWLIYSPDQVHLLGKSPQAR 180

175 ISNPSQTTQAQMLLEASVSRSEGEIYYQYRAEDDTCCRADEITHHLOAQAQRYLHVY 234

181 ISNPLNVQTAQMLLEASVSRSEGEIYYQYRAEDACGCTDELAHPSATVQRYLVQVHY 240

235 GNTASTETPLDGSAPSOADWLFYLPDYGERSNNLKTTPAPSTTGSWLCRQDRFRYE 294

241 GNLTASDVFTPLNGDDPLKSGWMMFCLVFDYGERKNSLSEMPFLFKATGNLWLCRQDRFRYE 300

Sequence 1640, Ap  
Sequence 672, App  
Sequence 671, App  
Sequence 670, App  
Sequence 10, Appli  
Sequence 8, Appli  
Sequence 4, Appli  
Sequence 2, Appli  
Sequence 1016, Ap  
Sequence 1076, Ap  
Sequence 2, Appli  
Sequence 6, Appli  
Sequence 694, App  
Sequence 58, Appli  
Sequence 4, Appli  
Sequence 684, App  
Sequence 78, Appli  
Sequence 3892, Ap  
Sequence 96, Appli  
Sequence 27, Appli

```
QY 295 YGFEIRTRLCROVLAMVHHLQALDSKITEHNGPTLVSRLLNLYDESAIASTLTIVFVRVGH 354
Db 301 YGFEIRTRLCROVLAMVHHLQALDSKITEHNGPTLVSRLLNLYDESAIASTLTIVFVRVGH 360
QY 355 EQDGNVTVLPPLELAYQDFSPRIHAIHQMDVLANFNAIQRMQLVDLKGEGLLGLLYQDK 414
Db 361 EDNNTVTALPPLELAYQDFPEQETALWQSDVLANFNTIQRMQLLDLKGEGVPGIYQDR 420
QY 415 GAWWYSAORLGBIGSDATWEXKQPLSVIPSLQSNASLVINDINGDQOLDWVITPGGLRGY 474
Db 421 NGWYYSARQAGEENNAVWGMQQLPITPAVDQNASLMDINDGQOLDWVITPGGLRGY 480
QY 475 HSORPDSWTRPTPLNALPVEYTHPRAQLADLMGAGLSDLVLGPKSVRLYANTRGEPK 534
Db 481 HSORPDSWTRPTPLNALPVEYTHPRAQLADLMGAGLSDLVLGPKSVRLYANTRGEPK 540
QY 535 GKDVQSGDITLTPVGADPRKLVAFSDVLGSGQAHLEVSATKVTWPNLGRGRFQOPIIT 594
Db 541 GRDVQSGDITLTPVGADPRKLVAFSDVLGSGQAHLEVSATKVTWPNLGRGRFQOPIV 600
QY 595 LPQFSQATEFNPAQVILADLDSGPTDLIYVHTNRDLIFPKSGNGFAPBPVTLRPPGL 654
Db 601 LPQFSQASAFNDRVHLADLDSGPDLLIYVHADRLDIFSNESGNGFAKFPFLLSPDGL 660
QY 655 RPDHTCOLQWADVQGLGVASLILSVPHMSPHWRCDLTNMKPMLLNEMNNMVGHHTLRY 714
Db 661 RPDHTCOLQWADVQGLGVASLILSVPHMADPHWRCDLTNMKPMLLNEMNNMVGHHTLRY 720
QY 715 RSSQFWLDEKAAALATGOTFCVCLPFPPIHTLWQTEDEISGNKLVTLRYARGAWDGR 774
Db 721 RSSQFWLDEKAAALATGOTFCVCLPFPPIHTLWQTEDEISGNKLVTLRYARGAWDGR 780
QY 775 EREFRGFGYEQTDSHOLAQNAPERTPPALTKWNTATGLPVDNALSTBYWR-DQAFPA 833
Db 781 EREFRGFGYEQTDSHOLAQNAPERTPPALTKWNTATGLPVDNALSTBYWR-DQAFPA 840
QY 834 GPSPTTMDNDKVPITPDDNRSYWFNPAKGLLRSELYGLDSTNKHVPYVTTTEPR 893
Db 841 GPTPTFLWKEGKDVPTPDDNRSYWFNPAKGLLRSELYGLDSTNKHVPYVTTTEPR 900
QY 894 SQVRRLQHTDSRYFVLWSSVYSESNHYERIASDPQCSQMITLSSDRFGQPLKQLSVQYP 953
Db 901 PQVRQLQDGTATVPVLWASVYSESNHYERIASDPQCSQMITLSSDRFGQPLKQLSVQYP 960
QY 954 RROQPAINLYPDLTPDKLLANSYDDQORQLRLTYQSSMHLTNVTVRVLGLPDSRSDI 1013
Db 961 RRNKPTTNPYPDLTPDLTFASSYDDQOQLRLTCROSSMHLNGLNELRVLGLPDSRSDA 1020
QY 1014 FTYGAENVPAAGLNLLELSDKNSLIADDKPREVLGQKATYTDGONTTPTQALIA 1073
Db 1021 FTYDAQVPVDGVLNLELTCALNSLIADDKPREVLNQQRTYTDGOKQTPDKTPTQALIA 1080
QY 1074 FTETTVFNQSTLAFNGSPISDKLSTLLEQAGYQQTNYLFPRTGEDKVVVAHHGYTDYGT 1133
Db 1081 FTETAVLTESLLAFDGGITPDELPGLLTQAGYQQEPYLPFRGENKVVVARQGYTDYGT 1140
QY 1134 AAQFWRPQKQSTQTKITLINDANYCVVVQTRDAAGLTTSKAYDNWRFITPVLQTDIND 1193
Db 1141 EAQFWRPVPAQRNSILLTKMTLKWDTHYCVITQTDAAGLTVSANYDNWRFITPVLQTDIND 1200
QY 1194 NQHLITLDALGRPITLRFWCTENGKMTGYSSPEKASFSPSDVYNAATELKKPLVPAQCQV 1253
Db 1201 NVHLITLDALGRPVTQRFWGESVATGYSSSEKPEFPSPNDIDTALNGLPLVPAQCIV 1260
QY 1254 YAPESWMPVLQKTFNRLABQDMQKLYNARIITEDGRICTLAYRRWVQSKAIPQILSL 1313
Db 1261 YAPDSWMPVLSQETFNTLTQEEQETLDRSLIITEDWRIICALTRRWLQSKISTPLVKLL 1320
QY 1314 NNGRPLPPHSLTTLTDYHDPEQIQQQVVFSDGFORLLQAAARHBAGMARQNEGDGL 1373
Db 1321 TNSIGLPPHNLTLTDRYDRDSEQQIRQQVAFSDGFORLLQASVRHBAGMARQNEGDGL 1380
```

```
QY 1374 IINVOHTENRWAIVTGRTEYDNKGQPIRTYQPYFLNDRYVSNDSARQEKEADTHVYDP 1433
Db 1381 VTKVENTKRWAVTGRTEYDNKGQPIRTYQPYFLNDRYVSNDSAR--KEAYADTHVYDP 1438
QY 1434 IGREIKVITAKGWFRITLFTPMFTVNEDNDNTAAE 1468
Db 1439 IGREIRVITAKWLRQSQYFPWFVTVSEDNDDTAAD 1473

RESULT 2
US-11-113-424-53
; Sequence 53, Application US/11113424
; Publication No. US20050260713A1
; GENERAL INFORMATION:
; APPLICANT: Gangolli et al.
; TITLE OF INVENTION: Polypeptides and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-225
; CURRENT APPLICATION NUMBER: US/11/113,424
; CURRENT FILING DATE: 2005-04-21
; PRIOR APPLICATION NUMBER: 60/256,704
; PRIOR FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 60/311,590
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: 60/257,314
; PRIOR FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: 60/311,613
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: 60/315,617
; PRIOR FILING DATE: 2001-08-29
; PRIOR APPLICATION NUMBER: 60/307,506
; PRIOR FILING DATE: 2001-07-24
; PRIOR APPLICATION NUMBER: 60/322,358
; PRIOR FILING DATE: 2001-09-14
; PRIOR APPLICATION NUMBER: 60/294,075
; PRIOR FILING DATE: 2001-05-29
; PRIOR APPLICATION NUMBER: 60/288,153
; PRIOR FILING DATE: 2001-05-02
; NUMBER OF SEQ ID NOS: 190
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 53
; LENGTH: 2515
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-11-113-424-53

Query Match 1.8%; Score 139; DB 7; Length 2515;
Best Local Similarity 16.5%; Pred. No. 0.082;
Matches 236; Conservative 161; Mismatches 446; Indels 586; Gaps 59;

QY 1 MONSQDPSITELSLPKGGAGITGMGEALPTGP----DGMALSLPLPISAGRGYAPFT 56
Db 1027 VRDNDISOPELNWE----AVVSGRCLPGDEAHGCGDALAKADKLPKGIATSSDNI 1082

QY 57 LNTNSGA-----GNSPPGLGW-----DCNVMTIRRRTHFGVPHYDE 92
Db 1083 LYFADCTNIRWDRDGVSTLIGNHMKSHWKPIPCGTLKLEMHLPTELAVSPMDN 1142

QY 93 TDTFLG-----PGEVLVWADQPRDESTLOGINLGATFTVTGYRSRLSHESFRLYEW 144
Db 1143 TLHIIDHMLRMTDGRVRVISGRP-----LHCATASTAYDTDLATHTLV--- 1189

QY 145 QPKTKCTDFWLLIYSPDGQVHLLGKSPQARISNPSTQTTQTAQWLLASVSSRGEQIYYQY 204
Db 1190 MPQS-----TAFGPLGELY-VAESDSQRINKRVRVIGTDRITAPFAGAESKNCNL--- 1237

QY 205 RABDDTGCBADEITHLQATAQ-----RYLHIVYVYGNRTASSETLPGLDGSAPS 252
Db 1238 -----ERGCDFEAEHYLATSAKENTIAALAVTPDSHVHIADQANYRIRSVMSSIPEASPS 1293

QY 253 -----QADWLP-----YLVFDYGERSNLTKTPAPFTTGSWLC 285
Db 1294 REYEIYAPDMQOEIYIFNRFQGHVSTRNLTGETTYFTYVNTVNTSNGKLSLTVTDAAGNKVF 1353
```



```
Db 1290 -----GAVFLSDNSRRV--KIKSTVWVKDLVKNSEVAVAGTQDCLPDDTR 1335
Qy 469 PGLRGVHSQRPDGSWTRFTPLNALPVEYTHPRAQLADLACAGLSDLVLGPKSVRLYANT 528
Db 1336 CG-----DGG-----KATEATLTPRGTITVDKFL-----IY---1362
Qy 529 RDGFAGK---KDVVQSGDITLVPFGADPRKLVAFSDVLGSGQAHLVEVSATKVTCPNLLG 585
Db 1363 ---FVGTMIIRDQNGII-----STLLGSNDL-----TSARPLSC-----1395
Qy 586 RGRFGQITLPGSPATEFNPAQVYLAIDLGGSGPTDLIYVHTNRDLIDFLKNSGNGFAEP 645
Db 1396 -----DSVMDISQVRLEW-PTDLAINPDNS-----LYVLNN-----1427
Qy 646 VTLRPFEGRLFDHTCLOLMADVQGLVASLILSVPHMSPHHWRCDLTNMKPMLLNEMNN 705
Db 1428 VLLQISE-----NH-----QVRIVAG-----RPMH---CQVPGIDHLLSK-----1460
Qy 706 MGVHHTLRYRSSQFMDLDEKAAALTTGQTPVCYLPPFIHTLMOTET-EDSIGNKLVT- 763
Db 1461 VAHTAL-----ESATALAVSHNGVLI-----AETDEKKINRQVTTTS 1500
Qy 764 --LRYARGAWDGRERE-----FRG-FGYVEQTDSHQLAQGNAP-----798
Db 1501 GEISLVAGAPSGCDCKNDANCDCFGDGYAKD-----AKLNTPSLAVCADGELYVAD 1554
Qy 799 -----ERTPAL-TKNWYATGLPV-----IDNALSTEYWRDQAQAPGPRFTTWQ 843
Db 1555 LGMIRIFIRKPNFLNTQNMYSLSPIQOELYFDTTGKHYTQSLPTGDYLYNFYTTG 1614
Qy 844 DNKDVPLTPED-----DNSRYWF-----NRALK-----GOLLRSE 873
Db 1615 DG-DITLITDNGNWNVRDSTGCMPLVLVPGQVYVWTGNTSALKSVTTQGHELAMM 1673
Qy 874 LY-----GL-----DSDS-----TNKHVPY-TVTEFRSQVRRLOH-----TDSR 905
Db 1674 TYHGNSGLLATKNSNGWTTTFYEVDSFGRLTNTVFTPGTVSSFRSDTDSVHVQVETSSK 1733
Qy 906 -----YPLVSSVVESENHY-----ERIASBQC-----930
Db 1734 DVTITNLSASAGAFYTLQDV-----RNSYYIGADGSLRLLLANGMEVALQPHLLAGT 1790
Qy 931 -----SONITLSSDR-----PGOPLK-----QLSVQVPRRQOPA 959
Db 1791 VNTVGRKNTLPIDGLNLVHRQKEQAGQVTFGRRLRVHNRNLLSLDFDR-----1845
Qy 960 INLYPDTLPDKLLANSYDDQOR-QLRLTYQSSGHHHTNNNTVRLVGLPDSRSDIFTYGA 1018
Db 1846 -----VTRTEKI---YDDHRKFTLRILYDQAGRPSLWSPSSRLNGVN-----VTYS- 1888
Qy 1019 ENVPAG---GLNLELLSDK-----NSLIADDPREYLCQOKTAYTDGQNTTPIQT 1065
Db 1889 ---PGGYIAGIQRGIMSERMEYDQAGRITSRIADGKTWSYTYLEKSMVL-----LLH 1938
Qy 1066 PTRQALIAFTETTVFNOSTLSAFNGSIPSDKLSL--TLEQAGYQOYNLPRTG-----ED 1119
Db 1939 SORQYITEFDKNDRLSSVTM-----PNVARQLETIRSVGYIRNYIQPPEGNASVIQD 1991
Qy 1120 KVMVAHGYDY-GTAA-----QFWRPQKQNTQL-TGKITLMD-----ANY 1160
Db 1992 FTEDGHLHTFYLTGTRRVIYKYGKLSKLAETLYDTTKVSFTYDETAGMLKTINLQNEGF 2051
Qy 1161 CVVVQTRDAAGL-----TTSAKYDWRFLTPVQVLTNDNQHLLT-----LDAL 1203
Db 2052 TCTIRYRQIGFLIDRQIRFTEEGMVNARFDYNY-----DNSFRVTSMQAVINET 2101
Qy 1204 GRPITLRFWGTENKMTGYSSPERKASFPSPDVNAAI-----ELKKPLPVAQCQ 1252
Db 2102 PLPIDLYRDVDSKTSQPKFGVIYY---DINQILITTAVMHTTKHFDAYGRMEKVEQYE 2157
Qy 1253 VYAPESMPPVLSOKTFNRLABQD-----WOKLYNARIITEDGRICTLAYRRWVQ 1301
Db 1301 -----KIKSTVWVKDLVKNSEVAVAGTQDCLPDDTR 1335
```

```
Db 2158 IFRSLMYMTVQVQDNNMGRVVKELKVGVPYANTTRYSEYDA-----DGQLQTVSI-----2207
Qy 1302 SOKAIPOLISLLNNGPRL--PPHSLTLTTDRYDHDPEQOIRQOVVFSDFGRLQLQAAAR- 1358
Db 2208 NOKPLMRYSDYDLNGNHLHLLSPGNSAKLTPLRFD-----IRDRIT-----RLGDVQYKM 2255
Qy 1359 HEAGMARQORNE-----OSLIINVQHTENRWAVTORTBYDNKGQPIRT-----YQYPF 1406
Db 2256 DEDGFLRQGGDIFEYNSAGLLIKAYNRAGMSV--RYRYDGLGRVRSVSSKSHSHLQPF 2313
Qy 1407 LNDW-----RYVNSDSARQEKAYADT--HYV-----DPGREIKUITAK 1444
Db 2314 YADLTNPTKVTHLYNHSSEITSLYDYLQGLHFLAMELSSGDEFYIACDNIQTPLAVFSGT 2373
Qy 1445 G-WFRETLPMPFTVNEDEN 1463
Db 2374 GLMIKQILITAYGEIYMDTN 2393

RESULT 4
US-10-793-626-96
; Sequence 96, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PUB480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 96
; LENGTH: 1227
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-10-793-626-96

Query Match 1.5%; Score 119; DB 6; Length 1227;
Best Local Similarity 17.0%; Pred. No. 0.84;
Matches 211; Conservative 160; Mismatches 443; Indels 428; Gaps 59;

Qy 428 IGSADVWTKMQPLSVIPSLQSNASLVINDINGDQLDWITGPGRLGYHSQR-PDG---SW 483
Db 61 VKNGVITWENQ-----QIDYPCGPDMPPEFEPGRCPRGASFSW 98
Qy 484 TRFTPLNALPVEYTHPRAQLADLACAGLSDLVLIGPKSVRLYANTRDGFAKDKVVQSGD 543
Db 99 YEYSPLR---VKPYIRGKLLDLWTALKEE-----QKGNRIAMAS 136
Qy 544 IITLPVPGADPRKLVAFSDVLGSGQAHLVEVSATKVTCPNLCGRFRGFPQITLPGFSQAT 603
Db 137 IV-----ENEKAKQYKEARGKG---HVRAN---W-----KDAT 165
Qy 604 EFNPAQVYLAIDLGGSGPTDLIYVHTNRDLIDFLKNSGNGFAEPV---TLRPFGLRPDHT 659
Db 166 DIIAAQI-LYTIKKDGPDRIA-----GFT-PIPAMSMISYASGARFIN- 206
Qy 660 CQLQADVQGLVASLILSVPHMSPHHWRCDLTNMKPMLLNEMNNNVHHTLRYRSSQ 719
Db 207 -----LLGEMLSFYDWDYADLPASPQIWEQTD-----VPSS 240
Qy 720 FWLDEKAAALTTGQTPVCYLPPFIHTLMOTETDEISGNKLVTTLRYARGAWDGRERFR 779
Db 241 DWYNASVIMMGSNVPLTRTP-DAHFN---TEVRYKGAQVSVAP-----281
Qy 780 GFGYVEGTD--SHQLACGNAPERTPALTQNTWATGLFDNALSTEYWRD--DQAPAGF 835
Db 282 --DYAENVKPADHWLAPHPGTD-----AAVAQAMTHVILQYENQPNDMFINY 328
```







```

Db 1892 ---ORQIFEDM---WRSLAITWPSVARHTMQTIRSIGYRTNYNPNPESNASTIYDNE 1946
Qy 1019 ENVPAGLNLLELLSDKNSLIADDPREYLGQOKTAYTDGQNTTLPQTPTQALIAFTETT 1078
Db 1947 E-----GLLLQ-----TAPLG-----TSRRVLPKYRQTRLSLSEILYDSTR 1981
Qy 1079 V---FNOSTLSAFNGSIPSDKSLTLEQAGVQOQNYLFPRT-----GEDKWVAHGTGTDY 1131
Db 1982 VSFTYDETAGVLKTVNLQSDGFICTIR---YRQIGPLIDRQIFRSEDGMVNARFDYS-Y 2037
Qy 1132 GTAAQFWRPQKQSN-----TOLTGKI-----TLIMDANY-----CVVQTR-- 1167
Db 2038 DNSFRVTSMQVINETPLPIDLYQFDDISGKVEQFGFVGIYDINQIISTAVWYTKHP 2097
Qy 1168 DAAGLTTSKYD-WRFILTPVQLTDINDNQHLITL--DALGRPITLRFWGTENGKMTGYSS 1224
Db 2098 DAHGRIKEIQEIPRSL-----MYWITIOYDNMGR-VTKR-----EIKIGPPAN 2140
Qy 1225 PEKASFPSPDYNAAIEHKPLPVAQCVYAPESWMPVLSQKTFNRLAEOQWKLKNARI 1284
Db 2141 TTKYAYE--YVDGQQLQ-----TVYLNKIM-----WRYNYDL-- 2171
Qy 1285 ITEDGRICTLAYRRWQSQKAIPLISLNNPRLPPHSLTLTTDRYDHDPEQIQOQW 1344
Db 2172 ---NGN-----LHLN-----PSSARLTPLRYD----- 2192
Qy 1345 PSDGFGRLQAAR-HEAGMARQDNE-----GSLIINVQHTENRMAVTRTEYDNKQ 1397
Db 2193 LRDRITRLGQVQRLDEGFLRQGTETFEYSKGLLTVYVSKSGWTVIYR--YDGLGR 2250
Qy 1398 PIRT-----YQFPLNDWRY-----VSDASRQEKAYADT--HVY----- 1431
Db 2251 RVSSKTSIGLOHQLQFFYADLTPTTRITHVYNSHSSEITSLYYDLOGLHFAMEISSGDFYI 2310
Qy 1432 --DPIGREIKVITAKG-WFRETLTPWPTVNEDEN 1463
Db 2311 ASDNTGTPLAVFSSNGLMLKQOTAYGEIYFDSN 2345

```

RESULT 7

```

US-10-858-730-72
; Sequence 72, Application US/10858730
; Publication No. US20050255568A1
; GENERAL INFORMATION:
; APPLICANT: Bailey, Richard B.
; APPLICANT: Blomquist, Paul
; APPLICANT: Doten, Reed
; APPLICANT: Driggers, Edward M.
; APPLICANT: Madden, Kevin T.
; APPLICANT: O'Leary, Jessica
; APPLICANT: O'Toole, George
; APPLICANT: Trueheart, Joshua
; APPLICANT: Walbridge, Michael J.
; APPLICANT: Yorgey, Peter S.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR AMINO ACID
; FILE OF INVENTION: PRODUCTION
; FILE REFERENCE: 14184-030001
; CURRENT APPLICATION NUMBER: US/10/858,730
; PRIOR FILING DATE: 2004-06-01
; PRIOR APPLICATION NUMBER: US 60/475,000
; PRIOR FILING DATE: 2003-05-30
; PRIOR APPLICATION NUMBER: US 60/551,860
; PRIOR FILING DATE: 2004-03-10
; NUMBER OF SEQ ID NOS: 364
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 72
; LENGTH: 1192
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-10-858-730-72

```

Query Match

1.5%; Score 115.5; DB 6; Length 1192;

```

Best Local Similarity 21.5%; Pred. No. 1.5;
Matches 167; Conservative 84; Mismatches 288; Indels 237; Gaps 42;
Qy 293 YEXGFETIRRLCRQVLMYHHLQALDSKITERHGPT-----LVSRLL-----LNYD 338
Db 443 YEDGDGPESRFAKTMALVAEHAVALTIDEEQARTACKKVEIAERLINDITNGNGVD 502
Qy 339 ESAI-ASTLVFVRVHEQ---DGNVVTLPLLELAYQDFSPRHHAHQWPMV-----LAN 389
Db 503 ESSILDTLFTTATQGESRRDG---IETIE-AIRELKKRH-----PDVQTTGLSN 551
Qy 390 FNAIQRWQLVDLKGEGPLGLYODKGAWYRSQRIGESDAVTWEKMOPLSVIPSLQS 449
Db 552 ISF-----GLNPAARQVILNSVFLHCOEAG-LDSAI VHASKIILPMRIPREOR 598
Qy 450 NASLVDINGDQLDWITGPGRLGYHSQRDPGSWTRFTPLNALPVEYTHPRA-----QL 503
Db 599 NVA-----LDLV-----YDRRED-----YDLOELMRLFEVGSNAASSKEDRL 636
Qy 504 ADLMAGLSLDLVLIGPKSVPLIYANTRGCPAKGKDVQSGDITLTPVPGADPKLV----- 557
Db 637 AELAGLPLFE-----RLAQRIVDGERGLD-----ADLDEAMTQKPLQIINEHLLA 683
Qy 558 ---AFSDVLGSGOHL-----VEVSATKVTCPNLGRGFGQPIITLPGFSQAPATEFPA 608
Db 684 GKTIVGELFGSQQLPFVLOSAAVMAVA-----YLEPHMERSDDDSGKG 730
Qy 609 QVYLADLDGSGPTDLIYVHTNRDLDFLNKSGN-----GFAEPVTLRPEGLRFDHTCOLQ 663
Db 731 RIVLATVKG---DVHDIKGNLVDIILSNNGYEVVNIQIKQIATIL-----EVAEDK 779
Qy 664 MADVQGLGVLASLILSVPHMSPHWRCDLTNNKFWLLNEMNNMVGHTLRYRSSSQFWLD 723
Db 780 SADVVGMS-GLLVKSTVVMKEN-----LEEMNTR-GV-----AEKFPVL 816
Qy 724 EKAALTTGTQTPVCYLPFPFHTLWQTR---TEDEISGNKLVTLRYA-RG-----A 770
Db 817 LGGAALTRS-----YVENDLAEIYQGEVHVARDAFEGKLMDTMSAKRGEAPDENSPEA 871
Qy 771 WDGRERFRGFGYVEQTDSSHQ-----LAQGNA-----PERT-----PPALTKNW 809
Db 872 IKAREKE-----AERKARHQRSKRIAAQKAAEPEVEPERSDVAADIEVPAPFNGSR 925
Qy 810 YATGLPVID-----NALSTEYWRDQAFAGSPRPTTWQDNKDVPLTPBEDDNRSTWFN 862
Db 926 IVKGLAVADYTGILDERALFLGWLGRGQGGGPSY-----EDLVETEGRPLRYWLD 979
Qy 863 RALKGQLL--RSELYGLDDSTN-----KHVPYTVTFRSQVRRLQHTDSRYPVLWS 911
Db 980 RLSTDGILAHAAVYGYFPAVSEGNDIVLTPKPDAPVRYRFRHFPQQR--GRFLCI-A 1036
Qy 912 SVVESRNYHYERIASDPQCSQNTLSSDRFGQPLKLSVQYPRRQQPAINLYPDTL 967
Db 1037 DFIRSRELAARGEVDVLPQLVTM-----GQPIADFA-----NELFASNAYRDYL 1082

```

RESULT 8

```

US-10-485-517-145
; Sequence 145, Application US/10485517
; Publication No. US20050256299A1
; GENERAL INFORMATION:
; APPLICANT: University of Sheffield
; APPLICANT: Biosynexus Incorporated
; APPLICANT: Foster, Simon
; APPLICANT: Mond, James
; TITLE OF INVENTION: Antigenic Polypeptides
; FILE REFERENCE: P100629M0
; CURRENT APPLICATION NUMBER: US/10/485,517
; CURRENT FILING DATE: 2004-02-02
; PRIOR APPLICATION NUMBER: GB 0118825.9
; PRIOR FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: GB 02000349.9
; PRIOR FILING DATE: 2002-01-09

```



Db 2986 WYKGGSLPARHQTHGSQLRLHHVSPADSGEYVCRAGGPGPQEQASFTVTVPSSGSSY 3045  
Qy 686 HWRCDLTMKP-----WLL--MEMNNMGV----- 708  
Db 3046 RLRSVPISIDPPSTVQGGQDASPKLIHGAAPISLEWTRNQELEDNVHISPNISIT 3105  
Qy 709 -----HHTLYRSSQFWLDEKAAALTTGTPVCYLPFPHTLW-----QTETEIS 756  
Db 3106 IVGTRPSNHGTIRCASNAYGVAQSVNLSVHGPPPTSV-LPEGPVWVKVKAATLECVS 3164  
Qy 757 GNTKLVTLRYARGAWDREFFRGYVEQTDH-----QLAQGN 796  
Db 3165 AGEPRSSARWTRISSTPAKLEQRYGLM-----DSHTVLQISSAKPSDAGTVCLQAONALGT 3221  
Qy 797 APERTPPALTKNMYATGLPVI---DNALSTE---YMRDDQAPAGSPRFT-----T 841  
Db 3222 AQQVEIVDTGAMAPGAQVQAEAEALTVEAGHTATLRCSATGSPARTIHWSKLRSPLP 3281  
Qy 842 WQ-----DNKDVPILTPEDDNSRYWFRALKQGLLSRLYGLDDSTNKHV---PY--TYTE 891  
Db 3282 WQHRLEGDTLIIPRVAQDSSQYICNAT-----SPAGHAEATIIILHVESPPYATTVPE 3334  
Qy 892 FRS-----QVRRLQHTDSKYPVLWSSVVES-----RN--YHVERIASDP---QCS 931  
Db 3335 HASVQAGETVOLQCLAGTPTPLTFQNSRVGSSLPGRATARNELLHFERAAPEDSGRYRCR 3394  
Qy 932 -ONITLSSDRFGQPLKQ-----LSVQYPRRQOPAINLYPDLPDKLLANSYD----- 977  
Db 3395 VTNKVGSAAFAQLLVQGPGLSPATSIAGSTPTVQVTPQ-LFTKSGISAVEPHCAVPS 3453  
Qy 978 DOORQLRLTYQSSW-----HHLTNTVTVLGLPDSRSTRDIITY 1016  
Db 3454 DRGTQLR-----WPKEGGLPPEGHSVQDGLVRLIQNLQDSQCG---TY 3492

## RESULT 10

US-10-485-517-198  
; Sequence 198, Application US/10485517  
; Publication No. US20050256299A1  
; GENERAL INFORMATION:  
; APPLICANT: University of Sheffield  
; APPLICANT: Biosynexus Incorporated  
; APPLICANT: Foster, Simon  
; APPLICANT: Mond, James  
; TITLE OF INVENTION: Antigenic Polypeptides  
; FILE REFERENCE: P100629W0  
; CURRENT APPLICATION NUMBER: US/10/485,517  
; PRIOR FILING DATE: 2004-02-02  
; PRIOR FILING DATE: 2001-08-02  
; PRIOR FILING DATE: 2001-08-02  
; PRIOR FILING DATE: 2002-01-09  
; NUMBER OF SEQ ID NOS: 424  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 198  
; LENGTH: 805  
; TYPE: PRT  
; ORGANISM: Staphylococcus aureus  
US-10-485-517-198

Query Match 1.4%; Score 114; DB 6; Length 805;  
Best Local Similarity 18.2%; Pred. No. 1;  
Matches 158; Conservative 93; Mismatches 265; Indels 350; Gaps 42;  
Qy 842 WQDNKDVPILTPEDDNSRYW-FNRALK--GQLLSRLYGLD-----DSTNKHV-----PYTV 889  
Db 13 WQD-----YWDENKTFKNDNLGQKKFYALDMFPYPSGAGLVHGPGEYTA 58  
Qy 890 TEFRSQVRRLQHTDSKYPVLWSSVVESRNVHYRIASDPQCSQNTLSSDRFGQPLKQ-- 947  
Db 59 TDIISRYKRWQGVNVLPMGH-----DAFGLPAEQYA 90  
Qy 948 LSVQYPRRQOPAINLYPDLPDKLLANSYDQORQLRLT-----YQSSW----- 992

Db 91 LDTGNDPREFTKKYIOTFKQIKELGFSY--DWDRVNTTDPYKYKWTQWFIQIYNKGLA 149  
Qy 993 --HHLTNTVTVLGLPDSR---SDIFTYGAENV---PAGGLNLELLSDKNLSIAD--- 1040  
Db 150 YVDEVAVNWCPCALGTVLVSNEEVIDGVSERGHPYKPKMOWVLKITEYADQLLADLDDL 209  
Qy 1041 -----DKPREYLGQOKTA-----YTDGQ-----NTTPLQTFTRQALIAFTTETTVNQ 1082  
Db 210 DWPESLKDMQMNWIRSEGAKVDFDNTGKVEVFTTRPDTIYGASFLVSPHEALVNS 269  
Qy 1083 STLSAP-----NGSIPSDKLSITLQ-----AGYQOTVYLPRTGCD-KVWVAHH 1126  
Db 270 ITTDEYKEKVKAYOTEASKKSLERTDLAKSGVFTGAYATN---PLSKEKQVQIWLADY 326  
Qy 1127 GYTDYGTAAQFWRPQKQSNLTQAGKITLIWDANCYVVVQTRDAAGLTTSKAYDWRFLTPV 1186  
Db 327 VLSYTGTA-----IMAVPAHDDRDYEFACKFDLPILIEVI 361  
Qy 1187 QLTDI-----NNQHILIT-LD-----ALGRPITL----- 1209  
Db 362 ECGNVBEAAVTGEGKHINSGLDLENEAAITKAIQLLEQKGAGEKKNYKLRDLWLFSRQ 421  
Qy 1210 RFWGT-----ENGWMTGYSSBEKASFPSPD-----VNAAIEL-- 1242  
Db 422 RYNGEPPIVTHWEDGTWTTVPBELLPLLLPETDEIKPSGTGESPLANIDSFVNVVDEKGT 481  
Qy 1243 -----KKPLPVAQCQVYA-----PES---WMPV-----LS 1264  
Db 482 MGRRETNTPQWAGSCWYLYRVIDPKENKWLADPEKLKHLPLVDLYIGGVEHAVLHLY 541  
Qy 1265 QKTFNRL-----ABQDWQKLYNARIITDGRICCTLAYRWVQSQKAIPLISLNNG 1316  
Db 542 ARFWHKVLYDLAIVPTKEPQKLFNQGMILGEG-----NEKMSKSKGNVINPD 589  
Qy 1317 PRLPPLSLTITTDYDHDPEQIQQVVPFSDGGRLLQAAARHEAGHARQ--- 1367  
Db 590 DIVQSHG-----ADTLRLYEMF---WGPLDAAIAWSEKGLDGSRRFLDRVWRL 634  
Qy 1368 --NEDGSLIINVQHTENR-----WAVTGRTYDNKGQPIRTYQPYFLNDWRVVSNDSA 1418  
Db 635 MVNEDGTLSSKIYTTNKSLLDKYVNTQVTKVTE-----DFETLGNFTA 677  
Qy 1419 RQB-----KEAYADTHYDP-----IGREIKVITAKGWFRTTLFTPMFTV 1458  
Db 678 ISQLMWFINECYKVDEYKPYIEGVKMLAPIAPHIGEEL--WSKLGHESSITYPWPY 735  
Qy 1459 NE-----DENDTAAEVK-----KVQM 1474  
Db 736 DEALLVDDEVEIVVQVNGKLRKIKI 761

## RESULT 11

US-10-858-730-73  
; Sequence 73, Application US/10858730  
; Publication No. US20050255568A1  
; GENERAL INFORMATION:  
; APPLICANT: Bailey, Richard B.  
; APPLICANT: Blomquist, Paul  
; APPLICANT: Doten, Reed  
; APPLICANT: Driggers, Edward M.  
; APPLICANT: Madden, Kevin T.  
; APPLICANT: O'Leary, Jessica  
; APPLICANT: O'Toole, George  
; APPLICANT: Trueheart, Joshua  
; APPLICANT: Walbridge, Michael J.  
; APPLICANT: Yorgey, Peter S.  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR AMINO ACID  
; FILE REFERENCE: 14184-030001  
; CURRENT APPLICATION NUMBER: US/10/858,730  
; CURRENT FILING DATE: 2004-06-01  
; PRIOR APPLICATION NUMBER: US 60/475,000



APPLICANT: Zhang, Jie  
APPLICANT: Zhou, Ping  
APPLICANT: Cao, Yi-Cheng  
APPLICANT: Drmanac, Radjoe T.  
TITLE OF INVENTION: Novel Nucleic Acids and Polypeptides  
FILE REFERENCE: 785CIP4CN  
CURRENT APPLICATION NUMBER: US/11/000,463  
CURRENT FILING DATE: 2004-11-29  
PRIOR APPLICATION NUMBER: 10/291,265  
PRIOR FILING DATE: 2002-11-08  
PRIOR APPLICATION NUMBER: PCT/US01/02623  
PRIOR FILING DATE: 2001-01-25  
PRIOR APPLICATION NUMBER: 09/922,279  
PRIOR FILING DATE: 2001-08-03  
PRIOR APPLICATION NUMBER: 09/491,404  
PRIOR FILING DATE: 2000-01-25  
PRIOR APPLICATION NUMBER: 09/617,746  
PRIOR FILING DATE: 2000-07-17  
PRIOR APPLICATION NUMBER: 09/631,451  
PRIOR FILING DATE: 2000-08-03  
PRIOR APPLICATION NUMBER: 09/633,870  
PRIOR FILING DATE: 2000-09-15  
NUMBER OF SEQ ID NOS: 944  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 250  
LENGTH: 1907  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-11-000-463-250

Query Match 1.4%; Score 113; DB 7; Length 1907;  
Best Local Similarity 18.0%; Pred. No. 4.7;  
Matches 296; Conservative 181; Mismatches 530; Indels 640; Gaps 80;

17 GGGALITGMEA---LPTGPD-----GMAALS-----LPLISAGRYAPAFPLNYS 61  
589 GVGFTTIEARTAQSTPSAPPQKVMCSMGSTTVRVSWVPPPADSRNGVITQSVAYEA 648  
62 GAGN-----SPFLGWDNCVMTRIRRTTHFGVPHYDRTDFTLPGEGE-VLVVADQ- 109  
649 VDGEDRGHRVVDGISREHSSWD--LVGLEKTEYRVVWVRAHTDVGPGPSSPVLVRTDED 706  
110 -----PRESTLQGINLGNATFT-----VTGR-----SRLESHFSR----- 140  
707 VPSGPPRKVEEPLNSTAVHYVWKLVPVSKQHGQIRGYQTVYVRLNGEPGLPIIQDVM 766  
141 --LEYWQPKTKGTDKDWLYSPDQGVHLLGKSPQARISNPSQTTQTAQWLLLEASVSRGE 198  
767 LAEQWPEE-----SEDIETISGUTPBTYS-----VTVAAYTTKGD 805  
199 QIYYQYRAEDDTGCEADEITHLQATQRYLHIYVYGNRTASETLPG-----LD 247  
806 GARSKPKIVTTGAVPGRPMTMISTAMNTALLQWH---PPKELPGLLGYRLQYCREAD 861  
248 GSAPSQADWLPLYVDFDGERNNKLTTPAPFTTG-----SWLCRQDRFSR-----YEVGFE 298  
862 EAREN-----TIDFG-----KDDQHFTVTGLHKGTTIYFRLAAKNRAGLGESEFEKE 907  
299 IRT--RLRCQVLMYHHLQALDSKITEHNGPTLVSRLLIYVDESALASTLVFVRVREGHD 357  
908 IRTPELPSGPPQNLHVLTGTTSTTE-----LAWDPPVLA-----ERN 945  
358 GNVVTLPLELAYQDFSPRHHMQPMDVLNFAINQWQLVDLKGELPGLLYQDKGAW 417  
946 GRIIS---YTVVFRDINSQBE-----LQNIITDTRFTLTGLK-----PDTTYDIKVA 990  
418 WYRSQRGLGTGSDAATWKEWQPSVIPSLSQSNASLVVDINGDGLDWTITGFLRGVHSQ 477  
991 W-----TSKSGGPLS--PSIQS----- 1005  
478 RPDGSWTRFTPLNALPVEYTHPRAQLADLMGAGLSDLVLIGKPSVRLYANTRDGFAKGD 537  
1006 -----RTMPEQVEFAK-----NFRVAAMKTS 1027

538 VVQSGDITLVPVGADPRKLVAFSDVLGSGQAHILVEVSATKVTWPNLGRGRFGQPIITLPG 597  
1028 VLLSWEVPSYKSAVPPK-----ILYNGQS--VEVDGHSNR-----KLIA 1065  
598 FSPQATEFNPAQVYVLADLDCGSPDILYVHTNRLLDIFLNKSGNGPAPVTLR----- 649  
1066 DLQNTETYS-----FVLMNR-----GSSAGLQHLVSRITAPDLLPH 1102  
650 -----FPBGLRFDHTCOLQMAVDQGLGVASITLSVPHMSPHHWRCDLTNNKPKWL----- 698  
1103 KLPASAYIEDGRFD-----LSMPHVQD-----PSLVRFYIVV 1136  
699 --LNEMNNMGVHHTLYRSSQFMDLDEKAAALTTGOTPVCYLPFPPIHTLWQTEDEIS 756  
1137 VPIDRVGSGM--LTPRWSTPELELLEAEIQG-----GEEQRR 1175  
757 GNKLAVTLR--YARGAWD-----GREERFGRGVYVEQTDHSHQAQGNAPERTPPALTK 807  
1176 RRRQERLKPVAAQLDVLPEFTTLGDKKNYRGF-----YNRP--LSPDLSY 1220  
808 NYVATGLPVIDNALSTEYMRDDQAFAGFSRFTTQDNKDVPLTP-----EDNSRYM----- 860  
1221 QCFVLA-----SLKEPMDQRYAS-SP-----YSDEIVVQVTPAQOQSEPEMLWTGP 1267  
861 -----FNR-----ALKCOLLRS-----ELYGLDDST--N 882  
1268 VLAVALIILIVIAILLFKRKRTHSPSKDRQSGICGLKDSLLAHSDDPVEMRLNQTQPMR 1327  
883 KHPVYVTEFRSQVRRLQHTDSRYPLVWSSVVERVHYVERIASDQCSQNTILSSDRFG 942  
1328 DHPPIPTDLADNIERLUKANDGL-----KFSQYESI--DP-----GOOFT 1366  
943 QPLKQLSVQVPRRQOPAINLYP-----DTLP--DKLIANSYDQQRQLRLTYQOSS 991  
1367 WENSLEVNPKRYANVIAVDHSRVILTSIDVPGSDYINANVIDGRKQ----- 1417  
992 WHHLTNNVTVLGLPDSRSDIFTTYGAENVPAAGNLLELSDKNSLIADDKPRYLGOOK 1051  
1418 ----NAYIATQGPLPETMGDFWRVWWEQRTATVVMTRLEEKSRVKCDQ----- 1462  
1052 TAYTDGQNTTPLQPTTQALIAETETVFNQSTLSAFNGSIPSPDKLSTLEAQYQQTNY 1111  
1463 -----YMPARG-----TETGLIQVT-----LDTVELATYTVRTF 1493  
1112 LFPRTGDK-----VWVAHHGYTDYGT-AAQFWRPQKQSNTOQLTGKITLIWDANY 1160  
1494 ALHKSGSSEKRELQFOFMAW-PDHGVPEYPTPLAFLRRVKACNPLDAGPMVVCAGV 1552  
1161 ----CVVV-----QTRDAAGLTTSAKYDWRFLTPVQLTDINDNQHLLITDLALGR 1205  
1553 GRTGCFVIDAMLERMKHEKTVDIYGHVTCWRSQRYNM--VQTED---QVVFIEHALLE 1606  
1206 PITLRFWGTENGKMTGYSSPEKASFSPSDVNAIELKXPLPVAQCQVYAPESWMPVLSQ 1265  
1607 AATCGH--TEVPARNLYAHQKLGQVPPGSESVTAMELE----- 1642  
1266 KTFNRLAEQDWQKLYNARIITEDGRICTLAYRRVWQSKAIPOLISILNNGPRLPPhSLT 1325  
1643 --FKLLAS-----SKAHTSRFISAN-----LPCNFKKRLVNM-----PYELT 1679  
1326 LTTDRYDHDPEQIQQOVVFSDFGRLQLQAAARHEAGMARQORNBGSLIINVQHTENRWA 1385  
1680 RVCLQPIRGVEGSDYINASFLDGY-----RQKAVIATQ---GPL---AESTEDFW- 1724  
1386 VTGETEDYDKQPIRTTQPYFLNDWRVYVNSDSARQ-----EKEAYADTHVYDPIG----- 1435  
1725 ---RMLWEHNSTII-----VMLTKREMREKCHQYWPABRSARYQYFVVDPMABYNMPQ 1776  
1436 ---REIKVITAKGWFRRTL-----FTPW 1455  
1777 YILREFVTARDQSGSRTIIRQFQFTDW 1803



```
RESULT 14
US-10-647-956A-6
; Sequence 6, Application US/10647956A
; Publication No. US20050251878A1
; GENERAL INFORMATION:
; APPLICANT: Bowen, David
; APPLICANT: French-Constant, Richard
; APPLICANT: Rocheleau, Thomas
; APPLICANT: Waterfield, Nicholas
; TITLE OF INVENTION: DNA SEQUENCES FROM PHOTORHABDUS LUMINESCENS
; FILE REFERENCE: 61645
; CURRENT APPLICATION NUMBER: US/10/647,956A
; CURRENT FILING DATE: 2003-08-26
; PRIOR APPLICATION NUMBER: CURRENT APPLICATION NUMBER: US/09/817,514
; PRIOR FILING DATE: CURRENT FILING DATE: 2000-03-26
; PRIOR APPLICATION NUMBER: US 60/191806
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 6
; LENGTH: 915
; TYPE: PRT
; ORGANISM: Photorhabdus luminescens
US-10-647-956A-6

Query Match      1.4%; Score 112.5; DB 6; Length 915;
Best Local Similarity 18.8%; Pred. No. 1.6;
Matches 129; Conservative 85; Mismatches 222; Indels 249; Gaps 34;

QY 478 RPDGSTRFTPLNALPVEYTHRAQADLMGAGLSDVLIGPKSVRLYA-----NTRD 530
Db 328 RNDABATRFWRNKVAPENSYYTDSLQILSATGREMANIGQNNQLPSALPSDNNYTT 387
QY 531 GFAGKXDVQSGDIT-----LPVFGADPRKLVAESDVL 563
Db 388 NYTRSYSDHSGNLTQIRHSSPATONNYTVAITLSNRSGVSLTTLTDPNQV-----DTL 443
QY 564 GSQAHLVVSATKVTCPNLGRGFGQPIITLPGFSOPATEFPAPQVYLADLDGSGPTDL 623
Db 444 FDAGGHTSLPQGTLLW---TPRGELQVNNNGPG-----NEWRYDSNG----- 485
QY 624 IYVHTNRDLTFLNKGNGFAEPVTLRPFPEGLRFDHTCQLQMAQDVQGLGVASL----- 675
Db 486 ----MRQLKVSEQPTQNTTQQRVIYLP-GLEL-RTTQSNATTEELHVLITGEAGRAQV 539
QY 676 -ILSVPHMSPHHRCDLTNMKPMLNEMNNMGVHHTLRYR-----SSSQFWLDEKAAAL 729
Db 540 RVL-----HWE-----SGKP-----EDVNNQ-----LRYSDNLIQSQLELDNQGII 579
QY 730 TTGQTPVCYLPPIHTLMQFTEDEISGNKLVTLTRYARGAWDGRERFRG---FGY--- 783
Db 580 SEEE---YYPFGTALMAANSQTEAS---YKTIIRYS-----GKERDATGLYYGYTRY 626
QY 784 -----VEQTDSSHQAQGN-----APERTPPALTQKNYATGLPVID 818
Db 627 QPWAGRWLSADPAGTIDGLNLYRMVRNPNVSLQDENGLAPEKGYTKEVNFP----- 678
QY 819 NALSTEWWRDQAF--AGFSRFTTQDNKDQVPLTPEDDNSRYWFNRLK----- 866
Db 679 -----DELAKFLAAKSHVYKWNKEK-----SSYTKNKSLLKVVVRVGDSDPS 719
QY 867 GQLLRSE--LYGLDDSTNKHVPYTVTEFRSQVRRLQHTDSRYPVLWSSVVSERNHYERI 924
Db 720 GYLLSHEELKGLIE-----KSGI-----YYSRL 742
QY 925 ASDPQCSQ-----NITLSSDRFG---QPLKQLSVQYPRROQPAINLYPD-----TLPD 969
Db 743 EENSLSLEKSKTNLSLGSLSGYMARTIQDTISEYAEHKYRSN-HPDFYSETDFFPALMD 801
QY 970 KLLANSYDDOORQLRLTYQSSWHLLTNNTVRVLGLPDSTRSDIPTYGAENVAPAGLNLE 1029
Db 802 KSBKNDYSG-ERKIYAAMEVKVYHDLKNK-----QSELHVNVALAHPTYQLSNE 849
```

```
QY 1030 ---LLSDKNSLIADDKPREYLGQOK 1051
Db 850 ERALLQETEPATAIDREYNFKGVGK 874

RESULT 15
US-10-510-386-218
; Sequence 218, Application US/10510386
; Publication No. US20050244922A1
; GENERAL INFORMATION:
; APPLICANT: Andersen, Jens Tonne
; APPLICANT: Clausen, Ib Groth
; APPLICANT: Jorgensen, Steen Troels
; APPLICANT: Olsen, Peter Bjarke
; APPLICANT: Rasmussen, Michael Dolberg
; TITLE OF INVENTION: Improved Bacillus Host Cell
; FILE REFERENCE: 10294.204-US
; CURRENT APPLICATION NUMBER: US/10/510,386
; CURRENT FILING DATE: 2004-10-04
; NUMBER OF SEQ ID NOS: 248
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 218
; LENGTH: 1432
; TYPE: PRT
; ORGANISM: Bacillus licheniformis
US-10-510-386-218

Query Match      1.4%; Score 112.5; DB 6; Length 1432;
Best Local Similarity 20.1%; Pred. No. 3.3;
Matches 229; Conservative 147; Mismatches 408; Indels 357; Gaps 64;

QY 298 EIRTRLCRQVLMYHHLQALDSKITEH-NGPTIVSLILNYDESATASTLVFVRRVHEQ 356
Db 73 QVDTPKVQAKAEKNAKSLSAKTYKRSVVSALRVKADESQ-SDLKRYLKK--QEK 129
QY 357 DGNVVTLPLELAYQDFSPRHHAHMPMDVLANFNAI-----QRWLVDLKGEGPLGL 410
Db 130 QGDV---KKIRSYIYVNGMAVHATKEVMEQVAAFEVEKVLNPKKQLIKPTQSVKXSA 186
QY 411 YQD-KGAWM-----VRSQRLGEISQSDAVTWKMQPLSVIPLSQNASLVDINGDQLD 463
Db 187 AKDEKEIENINRVDAPKAKLGYDGGTGV-----ASI-----DTGVQ 225
QY 464 WVTITPGL---RGYHSQRPDGSWTRFTPLNAL-----PVEY-THPRAQLADLMGAG 510
Db 226 W--DHAPLAKKRYGYPQHPDQPNHFEFSWYDAVSGASBPDDLEHGHVHTGTWVGSEPDG 283
QY 511 LSPL-VLIGPKSVRLYANTRDGFAGKDVVQSGDITLPVFGAD--PRKLVAFSDVLG--- 564
Db 284 QNIGVAPGAKWIAVKAFFSDDG-GTDEDLIAAGEMILAPKDKGNPHPEMA-PDVVNSW 341
QY 565 SQAHLVVSATKVTQW-----PNLGRGR--FGQITLPGFSOPATEFPAPQVYLADLD 616
Db 342 SGGAGIDEFYRDIYVKAWRAAGIPPEFSAGNVDPANP-----GGPASIAPANP--YPEAF 393
QY 617 GSGPTDLIYVHTNRDLTFLNKGNGFAEPVTLRPFPEGLRFDHTCQLQMAQDVQGLGVASLI 676
Db 394 ATGATDI-----DNRLADF-----SLOGPS--PYDET-----KPEISAPGV-NIR 430
QY 677 LSVPHMSPHHRCDLTNMKPMLNEMNNMGVHHTLRYRSSSQFWLDEKAAALTTGTPV 736
Db 431 SSVPG-SGYQDGDGDTSM-----AGPHVAAAAALIKQADSSITVDETEKILMETATPL 482
QY 737 CYLPPPIHTLMQFTEDEISGNKLVTLTRYARGAWDGRERFRGFGYVEGTDSHQLAQ- 795
Db 483 TDSKF-----TESPNNGYGHGLVNVFDAVSATVDGLGR-----ABGQ 519
QY 796 ---NAPERTPPAL-----TKNWTATGLP-----VIDN-----ALSTEYWRDDQAFAGFSR 839
Db 520 VGKEGDKSPALNHOEITFIYSGSETPLKADVEDDVSIISVKLSYKTTDDADHETIAAQ 579
QY 840 TTWQDNKD-----VPLTPEDDNSRYWFNRLKQQLLRSELYGLDDSTNKHVPYTVTEFRS 894
```

```
Db 580 TSGDYKKGTYEAAVTSEKGLTSYKWIUTDFGGNKTESKYVEFIS-----PAVTGYKQ 634
Qy 895 QVRELQHTDSRYPVLWSVSVESRNYHYERIASDPQCSQNTLSSDRFGQPLKQLSVQYPR 954
Db 635 DFB-----NSASGWLTSGVKNS-----WER-GIFQSGPNAAS-----GKNVPATNLATGPY 679
Qy 955 RQQPAINLY--PDTLPKLLANSVDDQORQLRLTYQOSSWHHLTNNTVRLVGLPDSTRSD 1012
Db 680 ESSANMLLMPVSVSP-----KNQKLYLTYKY--WRDIEED----- 713
Qy 1013 IFTYG-----AENVPA-----GGLNLELLSDKNS----- 1036
Db 714 -FDYGFVVVQPEGKEWIPAAEYSGKTSWKQGGIDLSEYGGQTIKWFNQLQSDSIEGD 772
Qy 1037 -LIADD-----KPREYLGQOKTAYT 1055
Db 773 GLYIDDDVALVKEVKSAGTKKRLGVEKQAKMKDKKTKRMIDPKKAKPAEAL-QEK--T 828
Qy 1056 DGQNTTPIQTPTRQALLAFTET---TVFNQSTLSAFNGSIPSDKLSLTLEQAGYQOQNYL 1112
Db 829 ETKKAAPAVLFPVR-AQVSVLETGKSTYSNQAT-GAYSLAHAPGTYTLKAEAYGYESTAO 886
Qy 1113 FPRYGEDKVVVAHHGYTDYGTAAQFWRPQKQSNLTQGT-----KITLIWDA 1158
Db 887 V-KIESDKT-----TTADFVLKELKKGTT-LGTINKKKTGEPVRHAKLYIIVEDA 933
Qy 1159 NYCVVVQTRDAAGLTTSKAYDWMRELTVPVQLTDINDNQHLITLDA-----LGRPI 1207
Db 934 -AVKPVQTDGGSYSLTA-YEGSYTVKVSANGYSSFSFVLDKGDVSKDIDLDPFICYPG 991
Qy 1208 TL-----RFGWTENG---KMTGYSSPEKA-----SFSPPSDVNAAIE 1241
Db 992 EIGYDDGTGENAWAFYESGNGLAVKMTLENQEKAMKGLGKFWDTTEFPDPGGTDPFAVE 1051
Qy 1242 L 1242
Db 1052 V 1052
```

Search completed: December 16, 2005, 14:47:30  
Job time : 31 secs

***This page blank (uspto)***

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 16, 2005, 14:33:31 ; Search time 197 Seconds  
(without alignment)

3126.297 Million cell updates/sec

Title: US-10-706-424-10  
Perfect score: 7901  
Sequence: 1 MNSQDPSITELSLPKGGGA.....WFTVNDENDTAAEVKKVKM 1474

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA Main:

- 1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*
- 3: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*
- 4: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*
- 6: /cgn2\_6/ptodata/1/pubpaa/US11\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	7901	100.0	1474	4	US-10-706-424-10
2	7901	100.0	1474	4	US-10-754-115-45
3	6031.5	76.3	1476	3	US-09-817-514A-4
4	4581.5	58.0	1485	4	US-10-262-794A-32
5	4261	53.9	1506	4	US-10-754-115-49
6	4261	53.9	1506	6	US-11-020-848-2
7	3971	50.3	1428	4	US-10-754-115-60
8	3834	48.5	1493	4	US-10-753-901-18
9	3834	48.5	1493	4	US-10-754-115-18
10	3106.5	39.3	1444	4	US-10-609-113-11
11	3106.5	39.3	1444	4	US-10-754-115-40
12	1262	16.0	697	4	US-10-609-113-39
13	202	2.6	2386	4	US-10-156-761-7751
14	199	2.5	2060	4	US-10-381-596A-2
15	181.5	2.3	2364	4	US-10-156-761-7834
16	162.5	2.1	1250	4	US-10-156-761-7572
17	160.5	2.0	1385	4	US-10-282-122A-68242
18	160.5	2.0	1426	3	US-09-912-020-340
19	160.5	2.0	1426	4	US-10-282-122A-42617
20	160.5	2.0	1426	5	US-10-771-241-340
21	160.5	2.0	1551	4	US-10-437-963-181412
22	159.5	2.0	1329	5	US-10-450-763-56139
23	159.5	2.0	1329	5	US-10-450-763-58761
24	154	1.9	5215	3	US-09-861-289-2
25	154	1.9	5215	3	US-09-860-846-2
26	154	1.9	5215	3	US-09-988-384B-2
27	154	1.9	5215	3	US-09-836-821-2

ALIGNMENTS

RESULT 1

US-10-706-424-10  
; Sequence 10, Application US/10706424  
; Publication No. US20040103455A1  
; GENERAL INFORMATION:  
; APPLICANT: ffrench-Constant, Richard  
; APPLICANT: Waterfield, Nicholas  
; TITLE OF INVENTION: DNA Sequences from tcd Genomic Region of Photorhabdus luminescens  
; FILE REFERENCE: 62878  
; CURRENT APPLICATION NUMBER: US/10706,424  
; CURRENT FILING DATE: 2003-11-12  
; NUMBER OF SEQ ID NOS: 16  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 10  
; LENGTH: 1474  
; TYPE: PRT  
; ORGANISM: Photorhabdus luminescens  
; US-10-706-424-10

Query Match	100.0%;	Score 7901;	DB 4;	Length 1474;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 1474;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	MNSQDPSITELSLPKGGGAITGMEALTPGPDGMAALSPLPLPISAGRGYAPAFITLNN	60	Sequence 45, Appl
Db	1	MNSQDPSITELSLPKGGGAITGMEALTPGPDGMAALSPLPLPISAGRGYAPAFITLNN	60	Sequence 21, Appl
Qy	61	SGAGNSPFLGWDGNCVNTIRRRTHFGVPHYDETDITFLGPEGEVLVADQPRDESTLOGIN	120	Sequence 13, Appl
Db	61	SGAGNSPFLGWDGNCVNTIRRRTHFGVPHYDETDITFLGPEGEVLVADQPRDESTLOGIN	120	Sequence 14, Appl
Qy	121	LGATFTVTGYSRSLSHSFSLRYWQPKTKTDFWLIYSPDGQVHLKSPQARISNPSQ	180	Sequence 157, Appl
Db	121	LGATFTVTGYSRSLSHSFSLRYWQPKTKTDFWLIYSPDGQVHLKSPQARISNPSQ	180	Sequence 1323, Ap
Qy	181	TTTQAQWLLASVSSRGEIYYQVRAEDDTGCCADEITHLQATAQRYLHVYGNRTAS	240	Sequence 93, Appl
Db	181	TTTQAQWLLASVSSRGEIYYQVRAEDDTGCCADEITHLQATAQRYLHVYGNRTAS	240	Sequence 5292, Ap
Qy	241	ETLPGLDGSAPSQADWFLYLVDFYDYSRNNLKTTPAFSTTGSWLCRODRFSRYEYGF	300	Sequence 5336, A
Db	241	ETLPGLDGSAPSQADWFLYLVDFYDYSRNNLKTTPAFSTTGSWLCRODRFSRYEYGF	300	Sequence 13, Appl
Qy	301	TRRLCROVLMYHHLQALDSKITEHNGTFLVSRLLYNDESIASTLVFVRVGHQDQNV	360	Sequence 46664, A
Db	301	TRRLCROVLMYHHLQALDSKITEHNGTFLVSRLLYNDESIASTLVFVRVGHQDQNV	360	Sequence 56547, A
Qy	361	VTLPPLSLAYODFSRPHHAHQMDVLANFNAIQRWQLVDLKGSLPGLLYQDKGAWMYR	420	Sequence 10384, A
Db	361	VTLPPLSLAYODFSRPHHAHQMDVLANFNAIQRWQLVDLKGSLPGLLYQDKGAWMYR	420	Sequence 42731, A

QY	421	SAQRLEIGSDAVTWKMQPLSVIPSLQSNASLVINDINGDQLDQWVITGPGRLGRYHSORPD	480
Db	421	SAQRLEIGSDAVTWKMQPLSVIPSLQSNASLVINDINGDQLDQWVITGPGRLGRYHSORPD	480
QY	481	GSWTRFTPLNALPVEYTHPRAQLADLMGAGLSDLVLIGPKSVRLYANTRDGFAGKDVVQ	540
Db	481	GSWTRFTPLNALPVEYTHPRAQLADLMGAGLSDLVLIGPKSVRLYANTRDGFAGKDVVQ	540
QY	541	SGDITILPVGADPRKLVAFSDVLGSGQAHVLSATKVTCPNLRGRGRFQPIITLPGFSQ	600
Db	541	SGDITILPVGADPRKLVAFSDVLGSGQAHVLSATKVTCPNLRGRGRFQPIITLPGFSQ	600
QY	601	PATEPAPAQVYLADLQSGPTDLIYVHTNRLDIFLNKSGNGFAEPVTLRPEGLRDPDHTC	660
Db	601	PATEPAPAQVYLADLQSGPTDLIYVHTNRLDIFLNKSGNGFAEPVTLRPEGLRDPDHTC	660
QY	661	QLQWADVQGLGVASLILSVPHMSPHWRCDLTNMKFWLLNEMNNMGVHHTLYRSSQF	720
Db	661	QLQWADVQGLGVASLILSVPHMSPHWRCDLTNMKFWLLNEMNNMGVHHTLYRSSQF	720
QY	721	WLDEKAAALTGTGTPVCYLPPDHTLWQTEDEISGNKLVTLRYARGAWDGRERFRG	780
Db	721	WLDEKAAALTGTGTPVCYLPPDHTLWQTEDEISGNKLVTLRYARGAWDGRERFRG	780
QY	781	FGVVEQTDHQAQGNAPERTPALTKWYATGLPVIDNALSTYWRDDQAFAGFSRFT	840
Db	781	FGVVEQTDHQAQGNAPERTPALTKWYATGLPVIDNALSTYWRDDQAFAGFSRFT	840
QY	841	TWQNDKDVLPDTPDDNSRYWFRNALKGQLLSELYGLDDSTNKHVPYVTFEFSQVRRQL	900
Db	841	TWQNDKDVLPDTPDDNSRYWFRNALKGQLLSELYGLDDSTNKHVPYVTFEFSQVRRQL	900
QY	901	HTDSRYPVLSSVSVESNYHYERIASDPQCSQMITLSSDRFGQLKQLSVQYPRRQPAI	960
Db	901	HTDSRYPVLSSVSVESNYHYERIASDPQCSQMITLSSDRFGQLKQLSVQYPRRQPAI	960
QY	961	NLYPDTLPDKLLANSYDDQORQLRLTYQSSWHLTNNTVRLGLPDSRSDIPTTGAEN	1020
Db	961	NLYPDTLPDKLLANSYDDQORQLRLTYQSSWHLTNNTVRLGLPDSRSDIPTTGAEN	1020
QY	1021	VPAGLNLLELSDKNSLIADDKPREYLGQQTAYTQGNTPPQTRQALIAFTETTVF	1080
Db	1021	VPAGLNLLELSDKNSLIADDKPREYLGQQTAYTQGNTPPQTRQALIAFTETTVF	1080
QY	1081	NQSTLSAFNGSIIPSDKLSITLLEAGYQOQNYLFPRTGEDKVVVAHGYTDYGTAAQFWRP	1140
Db	1081	NQSTLSAFNGSIIPSDKLSITLLEAGYQOQNYLFPRTGEDKVVVAHGYTDYGTAAQFWRP	1140
QY	1141	QKQNTQLTGKIITLIWDANCVVVQTRDAAGLTTSKYDWRFLTPVQLTDINDNQHILITL	1200
Db	1141	QKQNTQLTGKIITLIWDANCVVVQTRDAAGLTTSKYDWRFLTPVQLTDINDNQHILITL	1200
QY	1201	DALGRPITLRFWGTENGKMTGYSPEKASFPSPDVNAAIELKKPLPVAQCQVYAPESWM	1260
Db	1201	DALGRPITLRFWGTENGKMTGYSPEKASFPSPDVNAAIELKKPLPVAQCQVYAPESWM	1260
QY	1261	PVLSQKTFNRLAEDQWOKLYNARIITEDGRICTILAYRWVQSQAIPQLISLLNNGRPLP	1320
Db	1261	PVLSQKTFNRLAEDQWOKLYNARIITEDGRICTILAYRWVQSQAIPQLISLLNNGRPLP	1320
QY	1321	PHSLITLTDYDHDPEQIRQQVVFSDGFRLLQAAARHAGMARQORNEGSLIINVQHT	1380
Db	1321	PHSLITLTDYDHDPEQIRQQVVFSDGFRLLQAAARHAGMARQORNEGSLIINVQHT	1380
QY	1381	ENRWAVTGRTEYDNKGQPIRTYQPYFLNDWRYVSNDSARQEKAYADTHVYDPIGREIKV	1440
Db	1381	ENRWAVTGRTEYDNKGQPIRTYQPYFLNDWRYVSNDSARQEKAYADTHVYDPIGREIKV	1440
QY	1441	ITAKGWFRRLTLPWFVTVNEDENDTAAAEVKKVKM	1474
Db	1441	ITAKGWFRRLTLPWFVTVNEDENDTAAAEVKKVKM	1474

RESULT 2			
US-10-754-115-45			
; Sequence 45, Application US/10754115			
; Publication No. US20040208907A1			
; GENERAL INFORMATION:			
; APPLICANT: Hey, Timothy			
; APPLICANT: Schleper, Amanda			
; APPLICANT: Bevan, Scott			
; APPLICANT: Bintrim, Scott			
; APPLICANT: Mitchell, Jon			
; APPLICANT: Li, Ze Sheng			
; APPLICANT: Ni, Weiting			
; APPLICANT: Zhu, Baolong			
; APPLICANT: Merlo, Don			
; APPLICANT: Apel-Birkhold, Patricia			
; APPLICANT: Meade, Thomas			
; TITLE OF INVENTION: Mixing and Matching TC Proteins for Pest Control			
; FILE REFERENCE: DAS-104XCI			
; CURRENT APPLICATION NUMBER: US/10/754,115			
; CURRENT FILING DATE: 2004-01-07			
; PRIOR APPLICATION NUMBER: US 60/441,723			
; PRIOR FILING DATE: 2003-01-21			
; NUMBER OF SEQ ID NOS: 64			
; SOFTWARE: PatentIn version 3.2			
; SEQ ID NO 45			
; LENGTH: 1474			
; TYPE: PRT			
; ORGANISM: Photorhabdus luminescens			
US-10-754-115-45			
Query Match 100.0%; Score 7901; DB 4; Length 1474;			
Best Local Similarity 100.0%; Pred. No. 0;			
Matches 1474; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
QY	1	MQNSQDSITELSLPKGGGAIITGMGEALTPTGPDGMAALSLPLPISAGRYAPAFTLNYN	60
Db	1	MQNSQDSITELSLPKGGGAIITGMGEALTPTGPDGMAALSLPLPISAGRYAPAFTLNYN	60
QY	61	SGAGNSPFGLDGWCNVTIRRRTHFOVPHYDETDFTLGPGEVYLVADQPRDESTLQGIN	120
Db	61	SGAGNSPFGLDGWCNVTIRRRTHFOVPHYDETDFTLGPGEVYLVADQPRDESTLQGIN	120
QY	121	LGATFTVTGYSRLESHFSRLRYWQPKTKTDFTMLIYSPDGOVHLGLKSPQARIENPSQ	180
Db	121	LGATFTVTGYSRLESHFSRLRYWQPKTKTDFTMLIYSPDGOVHLGLKSPQARIENPSQ	180
QY	181	TTQTQWMLLEASVSSRGEQIYYQYRAEDDTGCEADBITHHLQATAQRYLHVYVGNRTAS	240
Db	181	TTQTQWMLLEASVSSRGEQIYYQYRAEDDTGCEADBITHHLQATAQRYLHVYVGNRTAS	240
QY	241	ETLPGLDGSAQSDWLFVLVFDYGRSNNLKTTPPAFTTGSWLCRQDRFSRYEYGFPEIR	300
Db	241	ETLPGLDGSAQSDWLFVLVFDYGRSNNLKTTPPAFTTGSWLCRQDRFSRYEYGFPEIR	300
QY	301	TRRLCQVLMYHHLQALDSKI TEHNGPTLVSRLLIYNDESAIASTLVFVRRVGHQDGNV	360
Db	301	TRRLCQVLMYHHLQALDSKI TEHNGPTLVSRLLIYNDESAIASTLVFVRRVGHQDGNV	360
QY	361	VTLPPLLEYAQDPSPRHHAWQPMQDVLNPFNAIQRWQLVDLKGEGPLGLLYQDKGAWYR	420
Db	361	VTLPPLLEYAQDPSPRHHAWQPMQDVLNPFNAIQRWQLVDLKGEGPLGLLYQDKGAWYR	420
QY	421	SAQRLEIGSDAVTWKMQPLSVIPSLQSNASLVINDINGDQLDQWVITGPGRLGRYHSORPD	480
Db	421	SAQRLEIGSDAVTWKMQPLSVIPSLQSNASLVINDINGDQLDQWVITGPGRLGRYHSORPD	480
QY	481	GSWTRFTPLNALPVEYTHPRAQLADLMGAGLSDLVLIGPKSVRLYANTRDGFAGKDVVQ	540
Db	481	GSWTRFTPLNALPVEYTHPRAQLADLMGAGLSDLVLIGPKSVRLYANTRDGFAGKDVVQ	540
QY	541	SGDITILPVGADPRKLVAFSDVLGSGQAHVLSATKVTCPNLRGRGRFQPIITLPGFSQ	600
Db	541	SGDITILPVGADPRKLVAFSDVLGSGQAHVLSATKVTCPNLRGRGRFQPIITLPGFSQ	600

Db 541 S G D I T L P V G A D P R K L V A F S D V L G S G Q A H L V E V S A T K V T C W P N L G R G F G Q P I T L P G F S Q 600  
Qy 601 P A T E N P A Q V Y L A D L D G S G P T D L I Y V H T N R L D I F L N K S G N G F A B E V T L R P P E G L R F D H T C 660  
Db 601 P A T E N P A Q V Y L A D L D G S G P T D L I Y V H T N R L D I F L N K S G N G F A B E V T L R P P E G L R F D H T C 660  
Qy 661 Q L Q M A D V O G L G V A S I L S V P H M S P H W R C D L T N K P W L L N E M N N M G V H H T L R Y R S S Q F 720  
Db 661 Q L Q M A D V O G L G V A S I L S V P H M S P H W R C D L T N K P W L L N E M N N M G V H H T L R Y R S S Q F 720  
Qy 721 W L D E K A A A L T T G Q P V C Y L P P I P H T L M Q T E D E I S G N K L V T L R Y A R G A W D G R E R F R G 780  
Db 721 W L D E K A A A L T T G Q P V C Y L P P I P H T L M Q T E D E I S G N K L V T L R Y A R G A W D G R E R F R G 780  
Qy 781 F G Y E Q T D S H Q L A Q N A P E R T P P A L T K N Y A T G L P V I D N A L S T E Y R D D Q A F A G S P R F T 840  
Db 781 F G Y E Q T D S H Q L A Q N A P E R T P P A L T K N Y A T G L P V I D N A L S T E Y R D D Q A F A G S P R F T 840  
Qy 841 T W Q N K D V L P T P E D D N S Y W F N R A L K G Q L L R S E L Y G L D D S T N K H V P Y T V T F R S Q V R R L Q 900  
Db 841 T W Q N K D V L P T P E D D N S Y W F N R A L K G Q L L R S E L Y G L D D S T N K H V P Y T V T F R S Q V R R L Q 900  
Qy 901 H T D S R Y P V L G S V V E S R N Y H E R I A S D P C S Q N I T L S S D R F G Q P L K O L S V O Y P R Q Q P A I 960  
Db 901 H T D S R Y P V L G S V V E S R N Y H E R I A S D P C S Q N I T L S S D R F G Q P L K O L S V O Y P R Q Q P A I 960  
Qy 961 N L Y P D T L P D K L L A N S Y D Q Q R L R T Y Q O S S W H H L T N N T V R V L G L P D S T R S D I F T Y G A E N 1020  
Db 961 N L Y P D T L P D K L L A N S Y D Q Q R L R T Y Q O S S W H H L T N N T V R V L G L P D S T R S D I F T Y G A E N 1020  
Qy 1021 V P A G L N L E L L S D K N S L I A D D K P R E Y L G Q O K T A Y T D G Q N T P L Q T P T R Q A L I A F T E T T V P 1080  
Db 1021 V P A G L N L E L L S D K N S L I A D D K P R E Y L G Q O K T A Y T D G Q N T P L Q T P T R Q A L I A F T E T T V P 1080  
Qy 1081 N O S T L S A F N G S I P S D K L S T L L E Q A G Y Q O T N Y L P R T G E D K V V A H G Y T D Y G T A A Q F W R P 1140  
Db 1081 N O S T L S A F N G S I P S D K L S T L L E Q A G Y Q O T N Y L P R T G E D K V V A H G Y T D Y G T A A Q F W R P 1140  
Qy 1141 Q K S N T Q L T K I T L I W D A N Y C V V Q T R D A A G L T T S A K Y D M R F L T P V Q L T D I N D Q H L I T L 1200  
Db 1141 Q K S N T Q L T K I T L I W D A N Y C V V Q T R D A A G L T T S A K Y D M R F L T P V Q L T D I N D Q H L I T L 1200  
Qy 1201 D A L G R P I T L R F W G T E N G O M T G Y S P E K A S F P P S D V N A A I E L K K P L P V A Q C Q V A P E S W M 1260  
Db 1201 D A L G R P I T L R F W G T E N G O M T G Y S P E K A S F P P S D V N A A I E L K K P L P V A Q C Q V A P E S W M 1260  
Qy 1261 P V L S Q K T F N R L A E D Q M K L Y N A R I T T E D G R I C T L A Y R W V O S O K A I P O L I S I L A N G P R L P 1320  
Db 1261 P V L S Q K T F N R L A E D Q M K L Y N A R I T T E D G R I C T L A Y R W V O S O K A I P O L I S I L A N G P R L P 1320  
Qy 1321 P H S L T L T T D R Y D H D P E Q I R Q O V V P S D G F G R L L Q A A R H E A G M A R Q N E D S G L I I N V Q H T 1380  
Db 1321 P H S L T L T T D R Y D H D P E Q I R Q O V V P S D G F G R L L Q A A R H E A G M A R Q N E D S G L I I N V Q H T 1380  
Qy 1381 E N R W A V T G R T E Y D N K G Q P I R T Y Q P Y F L N D W R Y V N S D S A R Q E K E A Y A D T H V Y D P I G R E I K V 1440  
Db 1381 E N R W A V T G R T E Y D N K G Q P I R T Y Q P Y F L N D W R Y V N S D S A R Q E K E A Y A D T H V Y D P I G R E I K V 1440  
Qy 1441 I T A K G W F R T L T P M F T N E D E N D T A A E V K K Y K M 1474  
Db 1441 I T A K G W F R T L T P M F T N E D E N D T A A E V K K Y K M 1474

RESULT 3  
US-09-817-514A-4  
; Sequence 4, Application US/09817514A  
; Patent No. US20020078478A1  
; GENERAL INFORMATION:  
; APPLICANT: ffrench-Constant, Richard  
; APPLICANT: Bowen, David  
; APPLICANT: Rocheleau, Thomas  
; APPLICANT: Waterfield, Nicholas  
; TITLE OF INVENTION: DNA SEQUENCES FROM PHOTORHABDUS LUMINESCENS

; FILE REFERENCE: 61645  
; CURRENT APPLICATION NUMBER: US/09/817.514A  
; CURRENT FILING DATE: 2000-03-26  
; PRIOR APPLICATION NUMBER: US 60/191806  
; PRIOR FILING DATE: 2000-03-24  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 4  
; LENGTH: 1476  
; TYPE: PRT  
; ORGANISM: Photorhabdus luminescens  
US-09-817-514A-4  
  
Query Match 76.3%; Score 6031.5; DB 3; Length 1476;  
Best Local Similarity 75.9%; Pred. No. 0;  
Matches 1120; Conservative 122; Mismatches 224; Indels 9; Gaps 3;  
  
Qy 1 M Q N S Q D S I T E L S L P K G G G A I T G M G E A L T P T P G M A A L S L P L P I S A G R G Y A P A F T L N Y N 60  
Db 1 M Q N S Q T F S V T E L S L P K G G G A I T G M G E A L T P A G P G M A A L S L P L P I S A G R G Y A P S L T L N Y N 60  
  
Qy 61 S G A G N S P F G L G W D C N V M T I R R R T H F G V P H Y D E T D T F L G P E G E V L W A - - - - - D Q P R D E S 114  
Db 61 S G T G N S P F G L G W D C G W A I R R R T S T G V P N Y D E T D T F L G P E G E V L V A L N E A G A D I R S E S 120  
  
Qy 115 T L Q I N G A T T V T V T G Y S R L E S H F S R L E Y W Q P K T T G K T D F W L I Y S P D G Q V H L L G K S P Q A R 174  
Db 121 S L Q I N G L A T T V T C Y S R L E S H E N R L E Y W Q P T T G A T D F W L I Y S P D G Q V H L L G K N P Q A R 180  
  
Qy 175 I S N P S Q T T O T A Q W L L E A S V S R G E Q I Y Y V R A E D D T C E A D E I T H H L Q A T Q R Y L H I V Y Y 234  
Db 181 I S N P L A N Y N Q T A Q W L L E A S I S H S E Q I Y Y Q V R A E D A G C E T D E L A A H F S A T V Q R Y L Q T V H Y 240  
  
Qy 235 G N R T A S E T L P G L D G S A P S O A D W L F Y L V D Y G E R S N N L K T P P A P S T T G S W L C R O D R F S R Y E 294  
Db 241 G N L T A S D V P F T L N G D D P L K S G W M F C L V D Y G E R K N S L S E M P L F K A T G N W L C R K D R F S R Y E 300  
  
Qy 295 Y G F I R T R L C R Q V L Y H H L Q A L D S K I T E H N G P T L V S R L I N Y D E S A I A S T L V F V R R V G H 354  
Db 301 Y G F E L R T R L C R Q I L M F H R L Q T L S G Q A K D D E P A L V S R L I D Y D E N A M V S T L V S V R R V G H 360  
  
Qy 355 E Q D G N V T L P R L E A Y Q D F S P R H A H W O P D V L A N F A I Q R W L V D L K G E G L P G L I Y Q D K 414  
Db 361 E D N N T V T A L P L E A Y Q F F E P E Q T A L Q S M D V L A N F N T I Q R W L L D L K G E G V P G L I Y Q D R 420  
  
Qy 415 G A M Y R S A Q R L G E T G S D A V T W E R K Q P L S V I P S L Q S N A S L V D I N G D G L D W I T G P L R G Y 474  
Db 421 N G W Y R S A Q R A Q E M A V T G K M Q L L P I T P A V Q D N A S L M D I N G D G L D W I T G P L R G Y 480  
  
Qy 475 H S Q R P D S W T R F T P L N A L P V E Y T H P R A Q L A D L M G A G L S D L V L I G P K S V R L Y A N T R D G P A K 534  
Db 481 H S Q H P D S W T R F T P L H A L P I E Y S H P R A Q L A D L M G A G L S D L V L I G P K S V R L Y V N N R D G P T E 540  
  
Qy 535 G K D V V Q S G D I T L P V G A D P R K L V A F S D V L G S G Q A H L V E V S A T K V T C W P N L G R G F G Q P I T 594  
Db 541 G R D V V Q S G D I T L P F G A D A R K L V A F S D V L G S G Q A H L V E V S A T Q V T C W P N L G H G R F G Q P I V 600  
  
Qy 595 L P G F S Q A T E N P A Q V Y L A D L D G S G P T D L I Y V H T N R L D I P L N K S G N G F A B E V T L R P P E G L 654  
Db 601 L P G F S Q A A S F N P D R V H L A D L D G S G P A D L I Y V H A D R L D I F S N E S G N G F A K E F T L S F P D G L 660  
  
Qy 655 R F D H T C Q L Q M A D V O G L G V A S I L S V P H M S P H W R C D L T N K P W L L N E M N N M G V H H T L R Y 714  
Db 661 R F D H T C Q L Q V A D V O G L G V V S L I L S V P H M A P H H W R C D L T N A K P W L L S E T N N N N G A N H T L H Y 720  
  
Qy 715 R S S Q F W L D E K A A A L T T G Q P V C Y L P P I P H T L M Q T E D E I S G N K L V T L R Y A R G A W D G R 774  
Db 721 R S S V Q F W L D E K A A A L A T G Q T P V C Y L P P P V H T L W Q T E D E I S G N K L V T L R Y A H G A W D G R 780  
  
Qy 775 E R E P R G F Y E O T D S H O L A Q N A P E R T P P A L T K N Y A T G L P V I D N A L S T E Y R - D D Q A F A 833  
Db 781 E R E P R G F Y E Q T D S H Q L A Q N A P E R T P P A L T K S W Y A T G L P A V D N A L S A G Y W R G D K Q A F A 840









Qy	651	PEGRLPFDHTCOLQWADVQGLGVASLILSVPHMSPHWRCDLTNNKWPWLNEMNNNGVHH	710
Db	663	PEGVMFNTCOLQVSDIQGLGAASIVLTVPHMTPRHRYDFTNKCPWLLNVINNRGAET	722
Qy	711	TLVYRSSQFWLDEKAAALTGTGTPVCYLPFPPIHTLWOTETEDEISGNKLVTTLRYAGA	770
Db	723	TLFYRSSAQFWLDEKQIEBGLGPAASYLFPPIHLWRNEALDEITGNRLUTKVWNTAHGA	782
Qy	771	WDGRERFRGFGVVEQTDSDHQLAQGNAPERTP-----PALTKWYATGLPVIDNALSTEY	825
Db	783	WDGREREFCGFRVQLTDTDEFAKGTI- EKAPDENIYPSRSISWPAFTGLPEVDSQLPAEY	841
Qy	826	WR-DDQAFAGFSRPTTW-----QDNQVPLTPEDDONSRYWFNRAKLGQLLSELYGLD	878
Db	842	WRGDDQAFAGTFRFTRYEKGNAQEQDTPIKEPTETEYMLNRAMKQLLASEVYG-D	900
Qy	879	DSNKK-HVPYVTVEFRSOVERLQHTDSRYPVWLSSVVESENHYERIASDPQCSNITLS	937
Db	901	DKTEKAKIPVTVEARQVRLIPSNDEBAAPSSWTSIENRSYHERIYVDVPSCKQOVVLK	960
Qy	938	SDRFGQPLKOLSVOYPRQOPAINLYPDTLPDKLLANSYDDQOQRLRLTYQQSSWHLTN	997
Db	961	ADEYGFPLAKVDIAYPRNKPQNPYDSLPDTLFADSYDDQOQKLYLTQQQSYHYLTQ	1020
Qy	998	NTVRVLGLPSTRSDIFTY----GAENVPAAGLNLELLSDKNLSIADDKPREYLQOQKTA	1053
Db	1021	QDDWVLGLTDSRYSESVHYAQTDAQSDIPRAGLILEDLLKVDGLIGKDKTFYLGQORVA	1080
Qy	1054	YTDGQNTPLQTPTRQALIAFTETTVNQSTLSAFNGSIPSDKLSLTLEOAGYQOQNYLF	1113
Db	1081	YVGGD----AEKPTRQVRVAYTAAFPDDNALHAFDGVIAPELDTQOLLAGGY----LLV	1132
Qy	1114	PRTGE-----DKYVVAHGVTGYTAAQWRPQKQSNQTLTGKITLIWDANYCVWQTRD	1168
Db	1133	PQISDVAGSSEKVMARQGYTEYGSAAQFYRPIIQKSLTGTGYLTSMWTHYCVWVKTD	1192
Qy	1169	AAGLTTSKTYDWRFLTPVLQTLTDINDNQHLYTLDAIGRPITLRFWGTENGKMTGYSSPEKA	1228
Db	1193	GAGMTTQAKYDYRFLPQAULTDINDNQHIVTFNALGVQTSRSRWGTEHNGISGYSPESK	1252
Qy	1229	SFSPSPDVNAAI ELKKPLPVAQCOVYAPESMPVLQKTFN----RLAEOQWKLYNARI I	1285
Db	1253	PFTVPDTEKALALOPTIPVSQCNIYVPDPSWMLLPQSLTGOLKEGETLWNLHRAGVV	1312
Qy	1286	TEDGRICITLAYRRWVQSKAIPQ-----LISLNGRPLPHSLTLTITDRDYDHPHQOIRQ	1341
Db	1313	TEDEGLICELAYRRWIKKQATSSMAVVTLOQLAQTPRPQPHAMTITTTDRYDSDSQOQLRQ	1372
Qy	1342	QVPSDFGRLILQAAARHEAGMARNEDEGLIIN-----VOHTENRWAVTGRTEYDNK	1395
Db	1373	SIVLSDGFRVLQQAQHEAGAEQWRAEDGSLVDVNTGKFPVANTTTTRWAVSGRTEYDGG	1432
Qy	1396	GQPIRTYQPYFLNDRWVSNDARSQKEAYADTHVYDPIGREIKVIITAKGWFRTTLFTPW	1455
Db	1433	GQAIRAYLPYFLNDRWVSDDARSAD--DLVADTHFYDPLGREYQVKTAKGFWEHMPMPW	1490
Qy	1456	FTVNEDENDTAAEV 1469	
Db	1491	FVNEDENDTAAEL 1504	

## RESULT 7

US-10-754-115-60  
 ; Sequence 60, Application US/10754115  
 ; Publication No. US20040208907A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Hey, Timothy  
 ; APPLICANT: Schleper, Amanda  
 ; APPLICANT: Bevan, Scott  
 ; APPLICANT: Bintrim, Scott  
 ; APPLICANT: Mitchell, Jon  
 ; APPLICANT: Li, Ze Sheng

```

; APPLICANT: Ni, Weiting
; APPLICANT: Zhu, Baolong
; APPLICANT: Merlo, Don
; APPLICANT: Apel-Birkhold, Patricia
; APPLICANT: Meade, Thomas
; TITLE OF INVENTION: Mixing and Matching TC Proteins for Pest Control
; FILE REFERENCE: DAS-104XC1
; CURRENT APPLICATION NUMBER: US/10/754,115
; CURRENT FILING DATE: 2004-01-07
; PRIOR APPLICATION NUMBER: US 60/441,723
; PRIOR FILING DATE: 2003-01-21
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: Patentin version 3.2
; SEQ ID NO 60
; LENGTH: 1428
; TYPE: PRT
; ORGANISM: Serratia entomophila
; US-10-754-115-60

```

Query Match	50.3%	Score 3971	DB 4	Length 1428		
Best Local Similarity	52.7%	Pred. No. 2.4e+304				
Matches	783	Conservative 186	Mismatches 431	Indels 86		
Gaps	16					
Qy	1	MONSDPSITELSI,PKGGGAI	TGGEAL	TTGPDGMAALSILPISAGRGVAPAF	LYN 60	
Db	1	MQHQDMAITAPTLP	SGGAVTGLGKDIAAAGDPGA	ATLSIPVSPGRGVAPT	GALNYH 60	
Qy	61	SGAGNSPFLGWD	CNVMTIRRRTHFGVPHYDE	TTDTFLGPEGEVLVA	-----DQPRDES 114	
Db	61	SRSGNGPFGIGWG	IGGNAVORRTENGAPT	YDDTDEFTGPDGEVLV	PALTAAGTQEARQAT 120	
Qy	115	TLOGINLGATFTVT	GRSRULESHFSRL	EYWPKTGKTDFWL	YSPDQVHLLGKSPQAR 174	
Db	121	SLLGINPGGSFNQ	VYRSRTEGSLR	LERLWLPADETET	EFWLYTDPDQVALLGRNAQAR 180	
Qy	175	ISNPSQTTQAWL	LEASVSSRGEQI	YQYRAEDDTGCE	ADEITHHLOATAQRYLHVY 234	
Db	181	ISNPTAPTQAV	MLMESSVLTG	EQMYQYQYRAEDD	CGDCBAERDAHPQAGAQRYPVAVWY 240	
Qy	235	GNRTASBTL	PCLDGSAPSQADW	LYLVFDYGERSN	NLKTTPAFSTTGS--WLCRQDRFSR 292	
Db	241	GNRQARTL	PAL--VSTPMSD	SWFLYLVFDYGER	SSVLEAPAWQTPGSGEWLCHQDCFSG 299	
Qy	293	YEXGFEIRRL	CRQVLMYHHLQAL	DSKITTEHNGPT	LVRSRLIILNYDSAJASTLVFVRRV 352	
Db	300	YEFGNLRL	RRLCRQVLMFYL	GLAGSSGANDAPAL	ISRLLLDYRESSPSLLENHVQV 359	
Qy	353	GHODGNVVTLP	PLELAYQDPS	PRHHWQPMVLAN	FNPAIORWQLVDLKGEGPLGLLYQ 412	
Db	360	AYESDGTSCAL	PALALGQWTF	FTPTPLSAMQTRD	DMGKLSLLQPYQLVDLNGEYGVGLTYQ 419	
Qy	413	DKGAWVRS	AQRLGEIGSDA	VTEKMQPLSVIT	SLQSNASIVDINGDGLDWTITGCLR 472	
Db	420	DSGAWVTR	BPVROSGDDP	DVATWGAAALPT	MPALHNSGILLADNGRLEWVWTAFCVA 479	
Qy	473	GYHSQRDGS	WTRFTPLNALP	VEYTHPRAQLAD	LMGAGLSDLVLIGPKSVRLYANTRDGF 532	
Db	480	GWYDRT	TEGRDMLHFTPL	SALPVEYAHKAV	LADILGAGLITDMLIGPRSVRLYSGKNDGW 539	
Qy	533	AKGKDVQSGD	ITLPVPGADPR	KLVAFSDVLGSG	QAHLVEVSATKYVTCWPNLGRGRFQGP 592	
Db	540	NKGETVQO	TERLTLPVPGVD	PRTLVAFSDMAG	SQQHLTEVRANGVYVWPNLGHGRFQGP 599	
Qy	593	ITLPGFSQ	PATEFNPAQVYL	ADLDGSGPTDL	IYVHTVRLDITFLNKSNGGFAEPVTLRFFE 652	
Db	600	VNIPIGFSQ	VTTFNPDQIL	ADTDGSGTTDL	IYAMSRLIYFNQSGNYFABPHTLLLPK 659	
Qy	653	GLRFDHT	COLQMAVQGL	GVASLITLSPH	MSPHWRCDLTNKKPWLLENMNNNGVHRTL 712	
Db	660	GVRYDRT	CSLQVADIQGL	GVPSLLITPV	APHVCHLSADKEWLLNGMNNNGARHAL 719	
Qy	713	RYRSSSQ	FWLDEKAAAL	TTGQTPVCYL	LPFPFIHTLWQTE	TETDEISGNKLVTTLRYARGAWD 772

Db 720 HYRSSVQFWLDEKAEALAGSPACYLPFTLTLHLSRVSQDEITGNRLVSDVLYRHGVM 779  
Qy 773 GRREPRFGVQVQTSQHLA-CGNAPERTPPALTKNWYATGLPVIDNALSTEWYRDD-Q 830  
Db 780 QQREPRFGVQVQTSQHLA-CGNAPERTPPALTKNWYATGLPVIDNALSTEWYRDD-Q 839  
Qy 831 APAGFSRFTTWQNDKDVPLTPEDDNRKYFWFNALQKQLRSELYGLDDSTNKHVPYTVT 890  
Db 840 AFADPATRFTVGSQEDSQYTP-DDSKTFWQLQALGILLRSELYGADGSSQADIPYSVT 898  
Qy 891 EFRSQVRLQHTDSRYVPLVSSVVSRYNHYRIASDPQSQNITLSSDFGPKQLKLSV 950  
Db 899 ESRPQV-RLVYANGDYVPMVPMGAESRTSVYRTHNDPQQQQAALLSDEYGFPLRQVS 957  
Qy 951 QYPRQCPALNLYPDTLPDKLLANSYDDQQRQLRLTYQQSSWHLLTNNTVR--VLGLPDS 1008  
Db 958 NYPRRPPSANTPASPATLFPANSYDEQQQILRLGLQQSSAHLVSLSGHMLLGLA 1017  
Qy 1009 TRSDIFTYGAENVPAGGLNLELLSDKNLSIADDKPREYLGQKQYATYDQNTTPTLPQTPTR 1068  
Db 1018 SRDDVFTYSADNPEGGGLTLEHLLAPESLVSQVGLAGQQQVWYLDSDQVATVAAPPL 1077  
Qy 1069 QALIAFTETTVFNQSTLUSAFNGSIPSDKLTTLLEAGYQQTNYLP--RTGEDKVVVAHH 1126  
Db 1078 PPKVAFIETAVLDEGMVSSLAAYIVDEH---LEQAGYRQSGYLPFRGREAEQALMTQCQ 1133  
Qy 1127 GYTDYGTAAQFWRPQKQNTQLTGKITLIWDANYCVVVQTFDAGLTTSKYDWRELTVP 1186  
Db 1134 GYTYAGAHEFWLPLSPRDSNLTGVPVTVTRDAVQCVITQWDAAGIVTTADYDMRFPTVP 1193  
Qy 1187 QLTDINDNOHLITLIDALGRPITLRFMGTEGKMTGYSSPKASFPSPSDVNAALIEKKPL 1246  
Db 1194 RVTDPNDNLQSVILDALGRVTLRFMGTEGKMTGYSSPKASFPSPSDVNAALIEKKPL 1250  
Qy 1247 PVAQCQVAPESMNPVLSQKTFNRLAEQDWOKLYNARIITEDGRICITLAYERVWQSKAI 1306  
Db 1251 PVAQCLVYVTDW-----GDDNDEK----- 1270  
Qy 1307 POLISLANNPRLPHSLTLTDRYDHPQOITRQOVVFSDFGRLLOAAARHEAGMARQ 1366  
Db 1271 -----MPPHVVLATDRYSDTQOQVRQVTFSDGFGRELSQATROAGNAWQ 1318  
Qy 1367 RNEDGSLI-----INVQTEENRWATGRTEYDNKGQPIRTQPYPLNDWRVYVND SAR 1419  
Db 1319 RGRDGLVATSDGLPVTVA-TNFRWATVGRAYDNKGLPVRVQPYPLDWSQVSDSAR 1377  
Qy 1420 QEKAVADTHYDPIGRKIVITAKGFWRTLTLPFWPTVNEDENDT 1465  
Db 1378 Q--DLVADTHYDPTAREWQVITAKGERRQVLYTTPWFVVSSEDENDT 1421

## RESULT 8

US-10-753-901-18  
; Sequence 18, Application US/10753901  
; Publication No. US20040194164A1  
; GENERAL INFORMATION:  
; APPLICANT: Bintrim, Scott  
; APPLICANT: Mitchell, Jon  
; APPLICANT: Larrinua, Ignacio  
; APPLICANT: Apel-Birkhold, Patricia  
; APPLICANT: Schafer, Barry  
; APPLICANT: Bavan, Scott  
; APPLICANT: Young, Scott  
; APPLICANT: Guo, Lining  
; TITLE OF INVENTION: Xenorhabdus TC Proteins and Genes for Pest Control  
; FILE REFERENCE: DAS-105X  
; CURRENT APPLICATION NUMBER: US/10/753,901  
; CURRENT FILING DATE: 2004-01-07  
; PRIOR APPLICATION NUMBER: US 60/441,717  
; PRIOR FILING DATE: 2003-01-21  
; NUMBER OF SEQ ID NOS: 22  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 18

; LENGTH: 1493  
; TYPE: PRT  
; ORGANISM: Xenorhabdus nematophilus  
US-10-753-901-18

Query Match 48.5%; Score 3834; DB 4; Length 1493;  
Best Local Similarity 50.9%; Pred. No. 1.8e-293;  
Matches 767; Conservative 213; Mismatches 472; Indels 56; Gaps 21;

Qy 1 MQNSQFSITELSLPGGGGAIKMGALFTGPDGMAALSPLPISAGRGYAFATLNYN 60  
Db 1 MQGSTPLKLEIPSPGGGSLKMGALNAVGAEGGASPLPISVGRGLVPLVSLNYS 60  
Qy 61 SGAGNSPPGLGWDNVTIRRRTHFGVPHVDETDPLGPEGL-VVAD---OP--RDES 114  
Db 61 STAGNSPFGMGWCGVGFSLRTAKGVPHYTGQDEYLGPDGEVLSPVDSQOQPEORTAT 120  
Qy 115 TLOGINLGATFTVTGYRSRLESFHSRLEYWQP--KTTGKTDFMLIYSPDQGVHLLGKSPQ 172  
Db 121 SLGLTVLTQPHVTTRYQSRVABKI VLEHWQPOQREETSFWLFTADGLVHFGKHH 180  
Qy 173 ARISNPSQTTQAWLLEASVSRGEQIYYQYRAEDDTGCEADEITHLQATQRYLHIV 232  
Db 181 ARIADPQDETRIARWLMEETVTHTGEHYHYHRAEDDLDCEHELAAQHSQVTAQRYLAKV 240  
Qy 233 YGNGRTASSETLPGLDGSAQADWLFVDFYDYGERSNNLKTPPAFSTTGS----- 282  
Db 241 SYGNTQPTAFPAVKSGIPADNDWLFHVFYDYGERSNSVSPFNVSENNVSENNVPEK 300  
Qy 283 WLCRQDRFSRYEYGFIRTRRLCRQVLMYHHLQALDSKITEHNGPTLVSRLLIINYDESAI 342  
Db 301 WRCRQDRFSRYEYGFIRTRRLCRQVLMYHHLQALDSKITEHNGPTLVSRLLIINYDESAI 360  
Qy 343 ASTLVFVRVGHQDQNVTLPLELAYQDFSPRHHQPMQDVLANFNAIQRWQLVDLK 402  
Db 361 VSLLOTARLAHETDGTVPVMSPLENDYQRVNHNLMQMSQPLEKMTLQPYQLVDLY 420  
Qy 403 GEGCLGLLYQD-KGAMWYSACR-LGEIGSDAVTWKMQPLSVIPLSQNASLVDINGDG 460  
Db 421 GEGISGLVQDQTKAWWYAPVRDITAEGTNAVYEAQPLPHIPAAQESAMLLDINGDG 480  
Qy 461 QLDWVITPGLRGYHSQRDPGSWTRFPLNALPVEYTHPRAQLADLMGAGSLDLVIGPK 520  
Db 481 RLDWVITASGLRGYHTMSPEGEWTPILPLSAVPMYFHPQAKLADIDGAGLPDLAGIPN 540  
Qy 521 SVRLYANTRDGFAKGKDVQSGDITLVPQADPKLVAFSDVLGSGOAHLEVSATKVTVC 580  
Db 541 SVRVSNRAGMDRAQDV:HLSDMPLFVPGNERHLVAFSDMTGSGQSHLVEVTADSVRY 600  
Qy 581 WPNLGRGRFGOPIITLPQSPQATEFNPAQVYLADLDGSGPTDLIYVHTNRLDIFLANKSGN 640  
Db 601 WPNLGHGKFGPELMTGFIQISGETFPNDRLYWVDIDSGTDDFIYARNYVLEIYANESGN 660  
Qy 641 GFAEPVTLRFPGLRFDHTCQLQADVQGLGVASLILSVPHMSPHWRCDLTNNKMWLN 700  
Db 661 HFAEPQRIIDLPGVRFDDTCRLQIATQGLGTASIIITLPHMKVQHWRLDMTIFKPLLN 720  
Qy 701 EMNNMGVHHTLYRSSSQFWLDEKAAALTGQTPVCYLPFPPIHTLWQTEDEISGNKL 760  
Db 721 AVNNMGTTETLYRSSAQFWLDEKLAQESGNTVSYLPPFVHVLMWRTVELDEISGNRL 780  
Qy 761 VTTLYARGAWDGRERERFGVVEQTDHQLA---QGNAPERTPALTKNWTATGLPVI 817  
Db 781 TSHYTHSHGAWDGLERERFGVVEQTDHQLA---QGNAPERTPALTKNWTATGLPVI 840  
Qy 818 DNALSTYWR-DDQAFAGFSPRETTWQDNK--DVLPTEPDNDSRYWFRNALKQLRSEL 874  
Db 841 DILLPTEYMGDQQA PPHFTFTRVDEKSGGDMTVP-SEQEYWLHRLAKQQLRSEL 899  
Qy 875 YGLDDSTNKHVPYTVTFEFSQVRLQHTDSRYVPLVSSVVSRYNHYRIASDPQSQNI 934  
Db 900 YGDDDSILAGTTPYSVDESRTQVRLPLVMVSDVPAVLVSVAESRQRYRYERVATDPQCSQKI 959

Qy 935 TSSDRFGQPLKQLSVQYPRQQAIPNLYPDTLPDKLLANSYDDQQQLRLTYQOQSSWHH 994  
Db 960 VLKSDALGFPQDNLEIAYSRPQPEFSPYDPTLFTLTSFDRQOMFLRLTRQSSYHH 1019  
Qy 995 LT-NNTVRLGLPDSRSDIPTYGAENVAPAGLNLLELLS--DKNSLIADDDKPREYLGQOK 1051  
Db 1020 LNHDNTWITGLMDSRSDARIYQADKVPDGGPSLEWFSATGAGALLLPDAAADYLGHQ 1079  
Qy 1052 TAYTDGQNTPLQPTTQALIAFTTTFVFNOSTLSAFNGSIPSDKSLTTLLEQAGYQOQNY 1111  
Db 1080 VAYTG-----PEEQAPLPLVAYIETABFDRSLAAPEEVWDEQELTKQLNDAGWNTAKV 1134  
Qy 1112 LPPTGDKWVAHGYTDYTAQAPRQKQSNQTLTKTLIWDANYCVVWOTRDAAG 1171  
Db 1135 PFSEKTDHVMVGQKEFTEYAGADGFRPLVQRETKUTGKTVTWDSHYCVITATEDAA 1194  
Qy 1172 LTTSKADWRFPLTPVQLTNDNDNHLITLDAALGRPITLRFWGTENGKWTGYSSPEKAS-- 1229  
Db 1195 LMQAHYDYRFWVADNTDNDNVHTVTFDALGRVTSFRFWGTENGKQGYTPAENETVP 1254  
Qy 1230 FSPSDVNAAILKPLPVAQCQYAPESMNPVLSQKTFNRLABQDWOKLYNARIITEDG 1289  
Db 1255 FIVPTTDDALALPGPIPVAGLMVYAPLSW---VQASFSNDGEL-YGELKPAGIITEDG 1310  
Qy 1290 RICTLAVRWQSQ--KAIQLISLANNPRLPPHSLTLTDRYDHDPEQOIRQOVVPSD 1347  
Db 1311 YLLSLAFRRQNNPAAAMPKQVNSQN-----PPHVLVSITDRYDADPEQQLRQTFTFS 1365  
Qy 1348 GFGRLLQAARHAGMARQNEGSLII-----NVQHTENRMAVTKRTEYDNKGQPIRT 1401  
Db 1366 FGRTLTQAVRHESGEAWRDEYGAIVAEHGAETAMTDFRMAVSGRTEYDVGKQALRK 1425  
Qy 1402 YQPYFLNDWRYVNSDSARKEAYADTHVYDPIGREIKVITAKGMFRTLFTPMFTVNE 1461  
Db 1426 YQPYFLNSQVNSDSSARQ--DIVADTHYDPLGREYQVITAKGGRFSLFTPMFVWNE 1483  
Qy 1462 ENDTAAEV 1469  
Db 1484 ENDTAGEM 1491

RESULT 9  
US-10-754-115-18  
; Sequence 18, Application US/10754115  
; Publication No. US20040208907A1  
; GENERAL INFORMATION:  
; APPLICANT: Hey, Timothy  
; APPLICANT: Schleper, Amanda  
; APPLICANT: Bevan, Scott  
; APPLICANT: Bintrim, Scott  
; APPLICANT: Mitchell, Jon  
; APPLICANT: Li, Ze Sheng  
; APPLICANT: Ni, Weiting  
; APPLICANT: Zhu, Baolong  
; APPLICANT: Merio, Don  
; APPLICANT: Apel-Birkhold, Patricia  
; APPLICANT: Meade, Thomas  
; TITLE OF INVENTION: Mixing and Matching TC Proteins for Pest Control  
; FILE REFERENCE: DAS-104XC1  
; CURRENT APPLICATION NUMBER: US/10/754,115  
; CURRENT FILING DATE: 2004-01-07  
; PRIOR FILING DATE: 2003-01-21  
; NUMBER OF SEQ ID NOS: 64  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 18  
; LENGTH: 1493  
; TYPE: PRT  
; ORGANISM: Xenorhabdus nematophilus  
US-10-754-115-18

Query Match 48.5%; Score 3834; DB 4; Length 1493;  
Best Local Similarity 50.9%; Pred. No. 1.8e-293;

Matches 767; Conservative 213; Mismatches 472; Indels 56; Gaps 21;  
Qy 1 MQNSQDFSIITELSLPQGGGAIITGMEALTTPDGMALSLPLPISAGRGVAPAFITLNYN 60  
Db 1 MCGSTPLKLBIPSLPSGGSLKMGBALNAVGAEGGASPSLPLPISVGRGLVPVLSLNY 60  
Qy 61 SGAGNSPFGGLGWCNVTIRRTTHFGVPHYDETDTLFLGPEGEVL-VVAD---QB--RDES 114  
Db 61 STAGNSFGMGWCGVGFISLRKAGVPHYTGDEYLGPDGEVLSIVPDSQGGQPEQRTAT 120  
Qy 115 TLOGINLGAFTVTGYSRSLSHFSRLEYWQP--KTTGKTDFWLIYSPDGOVHLLGKSPQ 172  
Db 121 SLILGTVLTPQHTVTRQSRVAEKIVLEHWQPOORREETSFWLFTADGLVHLFGKHH 180  
Qy 173 ARISNPSQTTQTAQWLEASVSSRGSIYYOYRAEDDTGCEADEITHHLQAQARYLHIV 232  
Db 181 ARIADPQDETRIARWLMEETVTHTEGHIYYHYRAEDDDLDCEHELAQHSQVTAQRYLAKV 240  
Qy 233 YGNRTASETLPGLDGSAQADWLFYLVFDYGRSNNLKTTPAFSTTGS----- 282  
Db 241 SYGNTQPETAPFAVKGIPADNDWLFHLVFDYGRSSSLNSVPFNFVSENNVSENNVPEK 300  
Qy 283 WLCRQRFPSRYEGFEIRTRRLCRQVLMYHHLQALDSKITEHNGTTLVSRLLINYDSAI 342  
Db 301 WRCPSPSFSRYEGFEIRTRRLCRQVLMFHLQALAGEKVAEETPALVSRLLILDYLNK 360  
Qy 343 ASTLVFVRRYGHQDQGNVVTLPFLEYAQDFSPRHHAHQPMQMDVLANFNIAQRQLVDLK 402  
Db 361 VSLQATRLAHETDGTPTVNMVSPLEMDYQRVNHGVNLNQMSQPMQKQNTLQPYQLVDLY 420  
Qy 403 GEGILPGLLIYQD-KGAWWYRSAQR-LGEIGSDATWKKQPLSVIPSLQSNASLVINDGDG 460  
Db 421 GEGISGLVYQDTQAWMYRAPVRDITAEAGTNAVYEEAKPLPHIPAQESAMLLDINDGD 480  
Qy 461 QLDWVITGCLRGVHSPDGRWTRPTPLNALPVEYTHPRAQADLMAGSLDLVLGPK 520  
Db 481 RLDWVITASGLRGVTHMSPEGEWTFPILSAVMEYFHPQAKLADIDAGLPLDALLPN 540  
Qy 521 SVRLYANTRDGFAGKQDVQVSGDITLPVPGADPRKLVAFSDVLGSGQAHLEVSATKVC 580  
Db 541 SVRWNSNNRAGWDRAQDVHLSDMWLPVPGNERHLVAFSDMTGSGQSHLVEVTADSVRY 600  
Qy 581 WPNLGRGRFCQPIITLPGFSQPAFEPNPAQVYVLADLDSGGTDLIYVHTNRLDIFLNKSN 640  
Db 601 WPNLGHGKFGLEPLMWTGFIQSGETFPNDRLYMYVDIDSGTTFDIYARNTYLEYANESGN 660  
Qy 641 GRAEPVTLRPEGLRPDHTCQLOMADVQGLGVASLILSVPHMSPHWRCDLTNNKPLLN 700  
Db 661 HFAEPQRIDLPDGVPRDDTCLRLQIADTQGLGTASIIITI PHMKVQHRLDWTIFKPMLLN 720  
Qy 701 EMNNMGVHHTLRYRSSSQFWLDEKAAALTGTGTPVCYLPFPIHTLMQTEDEISGNKL 760  
Db 721 AVNNMGTEITLYRRSSAQFWLDEKLAQSSSGMTVVSYLFPFVHVLRTEVLDISGNRL 780  
Qy 761 VTLIRYARGAWDGRERFRGFGVQDSDHQLA---QGNAPERTPPALTNNWYATGHPVI 817  
Db 781 TSHYHYSHGAWDGLERFRGFRVTQTDIDSRASATQCTHAEPAPSRRTNNWYGTGVREV 840  
Qy 818 DNALSTEYWR-DDQAFAFSPRFTTQDNK--DVPLTPEDDNSRYWFNRLKQGLLRSEL 874  
Db 841 DILLPTEYWGQDQAQFPHTFRFRTRYDEKSGGDMVTVP-SEQEYWLHRAKQGLRSEL 899  
Qy 875 YGLDDSTNKHVPYTVTFRQVRRLQHTDSRYVPLWSSVSVSESNRYHYERIASDPQCQNI 934  
Db 900 YGDDDSILACTPYSVDESRTQVALLPVMVSDPAVLVSAESRQYRVERVATDPCCQKI 959  
Qy 935 TLSSDRPGQPLKQLSVQYPRQQAIPNLYPDTLPDKLLANSYDDQQQLRLTYQOQSSWHH 994  
Db 960 VLKSDALGFPQDNLEIAYSRPQPEFSPYDPTLFTLTSFDRQOMFLRLTRQSSYHH 1019  
Qy 995 LT-NNTVRLGLPDSRSDIPTYGAENVAPAGLNLLELLS--DKNSLIADDDKPREYLGQOK 1051  
Db 1020 LNHDNTWITGLMDSRSDARIYQADKVPDGGPSLEWFSATGAGALLLPDAAADYLGHQ 1079





Db 1182 NNVAQVALDALGRVYVSRWTGTEBGIKTGP-RPE-VFEPATPETMEQALALASPLPVASCC 1239  
 Qy 1253 VYAPESMNPVLVSQKTFNRL---AEQDMQKLYNARIITEDGRICTLAYRRVWQSQKALPOL 1309  
 Db 1240 VYDAHSMNGTITTLAQSELVDPSEKQNSFLIDNRLIMPDGRIIRGRDRPWSLHLLPPAV 1299  
 Qy 1310 ISLIANGPRPLPHSLTITTTDRYDHPDPOQIRQOVVSGDGFGLLQAAARHAGMARQNE 1369  
 Db 1300 GELISEADRKPHTVILAAADRYPDPSQIQASIVFSDGFGRTIQTAKRE----- 1349  
 Qy 1370 DGSIIINVQHTENRWAVTGRTEYDNKGQPIRTYQPYFLNDWRYVNSDSARQEKAYADTH 1429  
 Db 1350 -----DTRWALAEVDYDGTGAVIRSFQPFYLLDDNNYGEAEV---SSSYATYI 1396  
 Qy 1430 VYDPIGRKIVITAKGFRRTLTFTWFTVNEDENT 1465  
 Db 1397 YYDALARQLRMVNAKGYERRTAFYFWPTVNEDENT 1432

RESULT 11  
 US-10-754-115-40  
 ; Sequence 40, Application US/10754115  
 ; Publication No. US20040208907A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Hey, Timothy  
 ; APPLICANT: Schleper, Amanda  
 ; APPLICANT: Bevan, Scott  
 ; APPLICANT: Bintrim, Scott  
 ; APPLICANT: Mitchell, Jon  
 ; APPLICANT: Li, Ze Sheng  
 ; APPLICANT: Ni, Weiting  
 ; APPLICANT: Zhu, Baolong  
 ; APPLICANT: Merlo, Don  
 ; APPLICANT: Apel-Birkhold, Patricia  
 ; APPLICANT: Meade, Thomas  
 ; TITLE OF INVENTION: Mixing and Matching TC Proteins for Pest Control  
 ; FILE REFERENCE: DAS-104XCI  
 ; CURRENT APPLICATION NUMBER: US/10/754.115  
 ; CURRENT FILING DATE: 2004-01-07  
 ; PRIOR APPLICATION NUMBER: US 60/441,723  
 ; PRIOR FILING DATE: 2003-01-21  
 ; NUMBER OF SEQ ID NOS: 64  
 ; SOFTWARE: PatentIn version 3.2  
 ; SEQ ID NO 40  
 ; LENGTH: 1444  
 ; TYPE: PRT  
 ; ORGANISM: Paenibacillus strain DAS1529  
 US-10-754-115-40

Query Match 39.3%; Score 3106.5; DB 4; Length 1444;  
 Best Local Similarity 43.0%; Pred. NO. 7.1e-236;  
 Matches 634; Conservative 230; Mismatches 551; Indels 61; Gaps 17;

Qy 3 NSQPSFTELSLPKGGGAIITGMEALPTGPDGMAALSPLPIAGRGYAPAFILNNSG 62  
 Db 5 SNADIKLSPSLPGKGGMKGIENIAAPGSDGMARCNVPLVTSGRYITPDISLYASG 64  
 Qy 63 AGNSPFLGWDGCVMTTIRRRTHFGVPHYDETDITFLGPEGVLVVADQPRDS-----STL 116  
 Db 65 HGNGAYGMGTGWGMSISRRISRTGTPSYTSDEQFLGPDGVLVPEESNEQGEIITRHTDTA 124  
 Qy 117 QGINLGATFTVTGYRSLRLESFSLRYWQPKTTGKT-DFWLIYSPDGQVHLGKSPQARI 175  
 Db 125 QGIPLGFTFTVTYFPRIESAFLLEYWEAQAGSATASFWLIHSADGVHLGCKGTAQARI 184  
 Qy 176 SNPGQTTQTAQWLLEASVSSRGEQIYYQYRABDDTGCEADEITHLQATQRYLHYVYG 235  
 Db 185 AAPDSDAKIAELWVEESVSPGEHIYYQYKEEDNQGVNLEEDNH--QYGANRYLKSIRYQ 242  
 Qy 236 NRTASETLPGLDGSAQADWLFLYVPDYGRSNNLKTTPAFSTTGWLCRODRFSRYEY 295  
 Db 243 NKVASPSLYVWKGELPADGQWLYSILDYIGENDTSAADVPLTTPQGEWLVPRDFRSYDY 302

Qy 296 GPEIRTRRLCROVLMYHHLQALDSKITERHGPTLVSRLLIILNYDESAIASTILVFRVRGHE 355  
 Db 303 GFEVRTCLRCQVLFHVFVKELGGE-----PALVWRMQLEYDENPAASMLSAVRQLAYE 356  
 Qy 356 QDGNVVTLPLELAYQDFSPRHEAHWQPMVDLANFNAIQRWQLVDLKGEGPLGLLYQDKG 415  
 Db 357 ADGAIRSLPPEPYTFPGIETTADWQFPLPVFEWADEEHYQLVDLYGEGIFGLLYQNND 416  
 Qy 416 ANWYRSQRIGETGSDAVTWKMQPLSVIPSLQSNASLVINDINGDQGLDWTGVTGFLRGYH 475  
 Db 417 HWHYRSPAR-GDT-PDGIAYNSWRPLPHIPVNSRNGMLMDLNGDYLEWLLAEFGVAGRY 474  
 Qy 476 SQPDGSGMTFTPLNALPVEYTHPRAQLADLMGAGLSDLVLIGPKSVRLVANTDGFAGK 535  
 Db 475 SMNPDKSWSGFVPLQALPTEFFHPQALANVTSGSLTDLVMI GPKSVRFVAGEBAGPKRA 534  
 Qy 536 KDVVQSGDITLVPFGADPRKLVAFSDVLGSGQAHLEVSATKVTCPNLGRGREGQPIITL 595  
 Db 535 CEVWQVGIITLPERVDKXELVAFSDMLGSGQSHLVRIHRDGVTCWNLGNVFGAPLAL 594  
 Qy 596 PGFSQPATEFPNPAQVYLADLDGSGPTDLIYVHTNRLDIFLNKSGNGPAEPVTLRFPBGLR 655  
 Db 595 HGFTASREFPNBRVYLVLDGSGASDIIVASRDALLIYRNLSGNGPADPVVPLPDGVR 654  
 Qy 656 PDHTCQLOMADVQGLGVASLILSVPHMSPHWRCDLTNMPWLLNENNNNGVHTLRYR 715  
 Db 655 PDNLCLLLPADIRGLGVASLVLPVYMPAPRSWKLLDFFAAKPYLLQTVNNLGAASSFWYR 714  
 Qy 716 SSSQFWLDEKAAALTTGQTPVCVLPPIHTLMQETETDEISGNKLVITTLRYARGAWDRE 775  
 Db 715 SSTQWLDKQAA-----SSAVSALPFPINVSMDHVTDEISGRTRTKYTYRHGVYRTE 770  
 Qy 776 REFRGFGYVEQTDSSHQAQGNAPERTPPALTKWYATGLPVIDNALSTYWRDDQAFAGF 835  
 Db 771 KEFAGFGRIDTWEEERDSEGLTSVSTPPLTRTYHTQKQDBERAVQYVWQGPAAQV 830  
 Qy 836 SP-RFTWQ--DNKDVLPTFEDDMSRYWFRNALXGQLLRSELVGLDLDSTKHVYTVTFE 892  
 Db 831 KPVRLTRFDAAAQDLFLDSNNNGQOEYLYRSRQGMPLRTEIFAGD--VGGSPPYQVESF 888  
 Qy 893 RSOVRLQHTDSRYPLVWSSVFSRNHYERIASDPOCSQNIILSSDFGQPLKQLSVQY 952  
 Db 889 RYQVRLVQSIDSECVL-PMQLEQLTYNYEQIASDPOCSQIQOWFDEYGVAAQSVTIQY 947  
 Qy 953 PRQQPAINLYPDTLPDKLLANSYDDQORLRLTYQSSSWHHLTNNTVRLVGLDPDSTRD 1012  
 Db 948 PRRAQPEDNPYRPLPDTSSSSYDSQMLRLTRQKAYHLADPEGWRLNI PHQTRLD 1007  
 Qy 1013 IFTYGAENVFAGGLNLELLSDKNSLIADDPREYLGQOKTAYTDGQNTTLPQTRQALI 1072  
 Db 1008 AFTYSADSVPAEGISABELLEVDTLRSSALEQAYGGQSEIYAGGGE-----PDLRALV 1061  
 Qy 1073 APTFTTVFNQSTLSAFNGSIPSDKLSTTLBOAGYQOQNTNLPFRPTGDEKVMVAHHGYTDY 1132  
 Db 1062 HYTRSAVLDSDCLQAYEGVSDSQNSLSSASSGYSRARSILGSGDEVDIFVAEQGFTRYA 1121  
 Qy 1133 TAAQWFRPQKQSTQTLTKITLIWDANYCVVQVTRDAAGLTTSKAYDWRFLTPVQLTDIN 1192  
 Db 1122 DEPNFFRILQQSSLLSSEQVLTWDDNFCAVTSIEDALGNQIQIAYDYRFEVIAQITDTN 1181  
 Qy 1193 DNQHLITLDALGRPITLRFMTGTENGKWTGYSSPEKASFSPSDVNAIIEKKPLPVACQC 1252  
 Db 1182 NNVAQVALDALGRVYVSRWTGTEBGIKTGP-RPE-VFEPATPETMEQALALASPLPVASCC 1239  
 Qy 1253 VYAPESMNPVLVSQKTFNRL---AEQDMQKLYNARIITEDGRICTLAYRRVWQSQKALPOL 1309  
 Db 1240 VYDAHSMNGTITTLAQSELVDPSEKQNSFLIDNRLIMPDGRIIRGRDRPWSLHLLPPAV 1299  
 Qy 1310 ISLIANGPRPLPHSLTITTTDRYDHPDPOQIRQOVVSGDGFGLLQAAARHAGMARQNE 1369  
 Db 1300 GELISEADRKPHTVILAAADRYPDPSQIQASIVFSDGFGRTIQTAKRE----- 1349  
 Qy 1370 DGSIIINVQHTENRWAVTGRTEYDNKGQPIRTYQPYFLNDWRYVNSDSARQEKAYADTH 1429

Db 1350 -----DTRMAIAERVDYDGTGAVIRSFQPFVLDWMVYGEAV--SSSMATIV 1396  
Qy 1430 VYDPIGREIKVIITAKWFRITLTPFWFTWNEDENDT 1465  
Db 1397 YYDALARQLRMVNAKGVERTAFTFPFWFTWNEDENDT 1432

## RESULT 12

US-10-609-113-39  
; Sequence 39, Application US/10609113  
; Publication No. US20040110184A1  
; GENERAL INFORMATION:  
; APPLICANT: Bintrim, Scott  
; APPLICANT: Bevan, Scott  
; APPLICANT: Zhu, Baolong  
; APPLICANT: Merlo, Donald J.  
; TITLE OF INVENTION: Pesticidally Active Proteins and Polynucleotides Obtainable from  
; TITLE OF INVENTION: Paenibacillus Species  
; FILE REFERENCE: DAS-101XC2  
; CURRENT APPLICATION NUMBER: US/10/609,113  
; CURRENT FILING DATE: 2003-06-27  
; PRIOR APPLICATION NUMBER: US 60/392,633  
; PRIOR FILING DATE: 2002-06-28  
; PRIOR APPLICATION NUMBER: US 60/441,647  
; PRIOR FILING DATE: 2003-01-21  
; NUMBER OF SEQ ID NOS: 49  
; SOFTWARE: Patentin version 3.2  
; SEQ ID NO 39  
; LENGTH: 697  
; TYPE: PRT  
; ORGANISM: Paenibacillus apairius strain DB482  
US-10-609-113-39

Query Match 16.0%; Score 1262; DB 4; Length 697;  
Best Local Similarity 37.0%; Pred. No. 2.5e-90;  
Matches 270; Conservative 119; Mismatches 291; Indels 50; Gaps 11;

Qy 736 VCYLPPPIHLMQTEDEISGNKLVITLAVAGAMDGRERFRGCVYEQTSHQLAQ 795  
Db 7 VCGVPPPIVVSIDHTVDEISGARIQKTYRNGVYDRTDKEFAGFHIDTWEERDSEG 66  
Qy 796 NAPERTPPALTKNWTGLPVIDNALSTYWRDQAPAGFSP-RFTWQ--DNKDVLPTP 852  
Db 67 TLSISPPVLTRTYHTGQDEERAVQVQWQDPAFAQVLPVLTFFDQATQDPLDS 126  
Qy 853 EDNRSRWFNRALQGQLRLRELY-----GLDDSTNKHVPYVTFEFSQVRLQHTDSRYP 907  
Db 127 PNRREYWLRLSLRGMPLRNEIFAGDVVGLP-----PYQVESLRYQVRLMQSTDSECV 179  
Qy 908 VLSVVESRNYHYERTASPOCSQNTLSSDRFGQPLKQLSVQYPRQOPAINLYPDTL 967  
Db 180 TL-PMOLEQTYNYEQIATSPQCSQQIQWDFEYGAQAQSIITQYPRRAQFEDNPVPHLL 238  
Qy 968 PDKLLANSYDDQORQLATYQQSSWHHLTNVTVLGLPSTRSIDFTYGAENVVPAAGLN 1027  
Db 239 PDTSWSSYDSQQMLLTLTRQKAYHLADPEGWELNIPHQTRLDSFIYADSVPAGIS 298  
Qy 1028 LELSDKNLSIADKPREYLGQQKATYDQNTTPTQTPTRQALIAFTETTVFNQSTLSA 1087  
Db 299 AELLGGDTLRSALPAEQYQGQSEIIVAGGE-----PDSRALVHYTRSAILDACLOA 352  
Qy 1088 FNGSPDCKLSTLEQAGYQQTNYLPRTCGEDKVVAAHGYTDYGTAAQFWRQKQNTQ 1147  
Db 353 YEGVLSQSLNSLASSGYQRSARILGSGDEADIFVAEQGFTRYADQNFRILGQOSSL 412  
Qy 1148 LTGKITLWIDANYCVVQTRDAAGLTTSKAYDNRFLTPVQLTDINDOHLITLDALGRPI 1207  
Db 413 LTGEQVLTWDDNFCVTSIEDALGNQIQIAYDRFVEAIIQITDANNVNVSLDALGRVV 472  
Qy 1208 TLRFWGTENGMTGYSSPEKASPPSDVNAALTELKPLPAQCVVYAPESWMPVLSQKT 1267  
Db 473 YSRWTGTEGIEGTGF-RPE-AEFSPPETMBEQALALASPLPVASCCVYDAHSWMGTITLQG 530

Qy 1268 FNRL---AEQDMQKLYNARIITEDGRICLTAYRRVQSQKAIPOLISLILNNGPRLPPHSL 1324  
Db 531 LSAIVPDSEKQMSFLIANELIMPDGRIARGRAPWMLQRLLPAPAVAKLSEADRKPEPHIV 590  
Qy 1325 TLTTRYDHDPEQIQIRQQVVFSDGFRLLQAAARHEAGMARQNEGSLIINVQHTENRW 1384  
Db 591 VLAADRYPDPSQIQASVVFSDGFGRTIQAKR-----ADTRW 629  
Qy 1385 AVGTREYDNKQOPINTYQPYFLNDWRVYNSDSARQEKAYADTHVYDPIGRKIVITAK 1444  
Db 630 AITERIDYDETGAIVRSFQFFYIDDMWYVYGEAV--SGSMYATIIYYIDALARQLRMVNAK 687  
Qy 1445 GWFRTLTTP 1454  
Db 688 GYERTATFYP 697

## RESULT 13

US-10-156-761-7751  
; Sequence 7751, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 7751  
; LENGTH: 2386  
; TYPE: PRT  
; ORGANISM: Streptomyces avermitilis  
US-10-156-761-7751

Query Match 2.6%; Score 202; DB 4; Length 2386;  
Best Local Similarity 18.3%; Pred. No. 1.7e-05;  
Matches 344; Conservative 184; Mismatches 647; Indels 706; Gaps 89;

Qy 13 SLPKGGGATG--MGEALPTG---PDGMAALSLPLPISAGR---GYAPAFPLTNYNS-- 61  
Db 318 TLASGSSDAGDYRASTLTSTGWEVSTGSGAFTYVNPQLPKPPMGSAPLSLSLNSQS 377  
Qy 62 -----GAGN---SPFGLGWDNCVMYTIIRR-----THFGVPH-----YDETDTFLGPEGEVL 104  
Db 378 VDGRTSASNQASWAGMGDLNVGIERRYVNCSEDLPTIGMCDWSPNSAKEPSGAVY 437  
Qy 105 VVADQPRDESTLQINGLATFTVTG---YRSRLSHFSRLYEQPKTKTKT-DFWLIYSP 160  
Db 438 VI-----NLNGVTSILQDNTGSGAYHLKNDPGWRVQRLPDGCGAGRDGSIYVISTQ 489  
Qy 161 DGOVHLGLKSPQAR-----ISN-----PSOTTQTAOWLLLEASVSSRG 197  
Db 490 DGRYFGWGRSERTATASVETEPVNGNDGTEPCDQFPPECTQAWRSLDRAVDANE 549  
Qy 198 EQIYYOYRAEDDTGCEADEITHLQATAOR-----YLHIVYVG----- 235  
Db 550 VETMYF-----DKEYNHVRSVANSDKAREVSSGVYKEIQYGWSSQIPDGKLP 599  
Qy 236 -----NRTASZ-----TLPLGLDGSAPQADWLPVLYFDYDYGSRNNL--KT- 273  
Db 600 KVLSHVNRCIERVBQBNPLRDEPATCPTFDKXTPSPVDPVDMCDGTSADYNCAGKTY 659



```
QY 254 ADWLFYLVDFYDGBRSNNLTPPAFTTSGWLCRODRFSRYEYFEIRTRRLCLQVLMYHH 313
Db 722 A-----LTGHELO--PAY-----SYIHTKDAFK-----HT 744
QY 314 LQALDSKITEHNGPTLVSRLL--ILNYDESA-----IASTLVFVRVVG-----HBQDGNVVT 362
Db 745 LTDVDVRNTN-----SGLGNVLNNEKANVIVKADTKDYMKYLSTDPDLNDSISVDK 797
QY 363 LPPLLEYAQDFSP-----RHAIHWQ-----PMDVLANFALQWQ 397
Db 798 LDHINHYLSLSAPDDTAIGNVRAMTNWLNELIRTSTVMNPSLYETTPSNVLKTRPVLNPIK 857
QY 398 LV-DLKGEGLPGLLYO-DKGAMWRS-----AORLGEIGSD-----431
Db 858 VIKXKNSGRITLYIYEAPEGYKWKNSQVKSITERSITQNTPEITFDIYNRGLTPTGTYS 917
QY 432 ---AVTWKMQPLSVIPSLQSNASLVINDINGDGLDWITGPGRLGYHSQRPDGSMTRFTP 488
Db 918 IRVATIWDENSEI-VRPTEBQSLS-----HNNLELSVYIT-EDLSGN-----KKFVS 962
QY 489 LNALPVEYTHPRAQLADL-MGAGLSDL-----VLIGPKSVLYANTRDGPAGKDVVQ 540
Db 963 VIDVPFKIALAKEYASTLTITGKDAASNSFKSQVDVNLG-ESVNLQNTAN-FTNSEGIK 1020
QY 541 SGDTILP-----VPGADPRKLVAFSD-----VLGSGQAHLEVESATKVTCPWN 583
Db 1021 EIIVTIPKONIKNTLALIPDBTKYRVVYTTDVRNGVNSNPTDLTKVAVYV-----1076
QY 584 LGRGRFGQP-----ITLPGFSQAPATEFNPAQVYLADLDG-----SGFTDLIYV 626
Db 1077 -----PDEPLVLTNGSQSFQNMVRVTPEDAPILTKAH-SQIFTKGLDNTWLSGNKVELET 1130
QY 627 HYNRLDI---FLKNSGNFABPVTLPPEGLRFDHTCQLQMAQVQGLGVASLILSVPHMS 683
Db 1131 EDNRGLVVKVYTNESGNTIQNSLTSGKKNTXN-----VSPQOM-1170
QY 684 PHWRCDLTNMKPWLLNEMNNMGVHHTLYRSSQFWLDEKAAALTGOTPCVCLPF--741
Db 1171 -----IDRLN-----RHYKFRVDNQ--LDPTGHYAKGQTKVNLVIYE 1208
QY 742 -----PIHLMQT-----ETDEISGNK-----759
Db 1209 VFGSVIADYKTTDGEVLSPLVTVNSQIBGTEYATPATIPDRVTFETTTDDGKVKKTIS 1268
QY 760 --LVTLRYARGAWDGRERFEGYVQDSDHQLAQGNAPERTPPAL-----805
Db 1269 YHLISTPENQSGTVGKQITIEVHYV--EPITTYEQIPNDAPQETPVALEVTRYVDSEGNE 1327
QY 806 -----TKNWTATGLPVIDNALSTE-YWRDDQAFAGFSPRET-----840
Db 1328 VQTEBEGTHDAGIADKMQYTQTAANGIITHVYQRIQSEIPNEAPQETPVALEVTCY 1387
QY 841 -----TWNQNDKVP-----LTPEDDNSRYFNWNRALKGQLLRSELYGLD 878
Db 1388 VDSEGEVQETEGTHDAPGIGDKMQYTQTTEDGITHIYQ-----IQSE---IP 1438
QY 879 DSTNKHVYVTVTFR-----SQVRLQHTDSRYPLVMSVBSRNYHYERIASDQCQONIT 935
Db 1439 NEAPQETPVALEVTRYVDSEGNEVQET-----BEGTHQPPSIIGDKMQYTQGT 1486
QY 936 LSBRFQQLKQLSVQYPRRQQAINLYPDTLPDKLLANSY-DDQQQLRLTYQSSWHH 994
Db 1487 TTADGI-----TYVVERIQSEIPNEAPKETPQLLEVTRYVDSEGNEVQET---EGGTHH 1538
QY 995 LTNNTVRVLGPDSTRSDIFTYGAENVYPAGLNLLELSDKNSLIADKPXEYLQ-QOKTA 1053
Db 1539 APG---IIG-----DKMQYTQTTESGIITHVYERIQSEIPNEAPQETPQLBVTR 1587
QY 1054 YTDQNTTPIQTP-----TRQALIAFTETTVFNQSTLSAFNQSIPSDKL 1097
Db 1588 YVNSEGEVQETEGTHQPPGIGDKMQYTQTTADGITYVYERIQSEIPNEAPKE-T 1646
QY 1098 STTLEQAGYQQTNYLPRPTGEBKVVVAHGYTDYGTAAQFWRPQKQSNQNTQLTGKITLWD 1157
```

```
Db 1647 PVQLEVTRYVDTD-----GNEVQETEBGTHQPPGLIGDKM-----QYTGRTV-----1688
QY 1158 ANYCVVVQTRDAAGTTSKDYMRFL-----TPVQL-----TPINDNQHLITLD 1201
Db 1689 -----EKDGITT---YVYERIQSAIPNEAPQETPVQLEVTRYVDITGNEVQETEE 1735
QY 1202 ALGRPITL-----RFWG---TENGMQTYSSPEKASPPSDVNAALIELKPLPVAQCQV 1253
Db 1736 GTHQRYIIGDKRWYSGVTVTENG-ITRHVY-ERIQSKVPNDA---POETPVQLEVTRY 1789
QY 1254 YAPESMMPVLSOKTFNRLAE-----NQIQTTEGKHQPPGIIGDRWQ---YTGKVTEDG-IITYVYER-I 1835
Db 1790 VDPEG-----NEIQETTEGKHQPPGIIGDRWQ---YTGKVTEDG-IITYVYER-I 1835
QY 1301 QSQKAIQILISLNGPRLPHSLTLTTDRYDHDPEQIQROQVVFSDGGRLLQAAARHE 1360
Db 1836 QSE--IP-----NNPQETPVELEVTR-----YVDGEGNEVQET-----1867
QY 1361 AGMARQORNEGSLINVOHTENRMAVTGR-TEVDNKGQPIRTYQVPYFLNDRVYVNSDSAR 1419
Db 1868 ---TEGKHQPPSI-----GDRMQYTGKVTEDG---ITTY--VYERIQSKVPNDAPR 1912
QY 1420 QEKEAYADTHVYDPIGRIKVTAKGWFRTLTFTWFTVNEDENDTAAEVKK 1471
Db 1913 VDIDELKITIYVDYTNGREI-VPSRKGQLPPEQF-----IGQDWQYTGKHIEK 1958

RESULT 15
US-10-156-761-7834
; Sequence 7834, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 7834
; LENGTH: 2364
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-7834

Query Match 2.3%; Score 181.5; DB 4; Length 2364;
Best Local Similarity 16.8%; Pred. No. 0.00072;
Matches 274; Conservative 180; Mismatches 460; Indels 715; Gaps 77;

QY 7 FSITSLSPKGGGAIITGMGEALTPTGP-----DGMALSLPLPISA-GRGYAPAFATLY 59
Db 305 YATVSSSSDAGDY---RASTLSPTGSWEVSTGSGSFNVSVFIQVPAPMGSAPLAMSY 361
QY 60 NS-----GAGN---SPFGLGWDGNVMTIRR---THFGVPH-----YDETDFTLGPE 100
Db 362 DSQSVDRGTSASNQASWSGMDLVNVTERRYNCTEDGLPTIGDMCWDSPNSAKEPD 421
QY 101 GEVLVADQPRBESTLQGINLGHATFTVTGYSRLESHPFSELYWQPKT-----TGKT-D 153
Db 422 GAAYVSLNGTTSSELIDNNNGSAY-----HLKNDPGRVQRULPDGHGANGGE 470
QY 154 FWLIYSPDGOVHLLGKSPQARISN-----PSQTTQTAQWLL 190
```

Db 471 YWVISTQDQRYFPGKRSERTSTATASVFTVPVGVNDAGEPCHDQPEPCTQAWRNLD 530  
Qy 191 ASVSSRGEQIYYQIRABDDTCCEADEITHHQAQRYLHVYVGNRTASTLPLGDSGA 250  
Db 531 RAVDANEVETWYFY-----DKEYNHYRSVA----- 555  
Qy 251 PSQADWLFYLVFDYGERSNNLKTTPAFSTTGSWLCRQDRFSRYEYGFIRRLCRQVLM 310  
Db 556 -----NTDKAREYSSGVYKE-----IQYGM----- 576  
Qy 311 YHHLQALDSKITEHNGPTLVSRLLINYDESIASTLVFRRVVGHEQDQGNVTLPLLELAY 370  
Db 577 -----SSQITDGKPAKVELSHVN-----RCIERVQENDPLARDEPAACPA 616  
Qy 371 QDPSPRHAWQPMVLANFNAIQRMQLVDLKGELFGLLYQDKGAWWYSAQBLGIGS 430  
Db 617 FDDKPGSYPD--VPVLDLMDGTSAD-----YNCAGKTYF----- 648  
Qy 431 DAVTWEKQPLSVTPSLQSNASLVYDINGDGLDWITGPGLRGYSQRPDGSWTRFTPLN 490  
Db 649 -----PTPFSTMDLMDI----- 660  
Qy 491 ALPVEYTHPRAQLADLMGAGLSLVLIGPKSVRLYANTRDGFAGKDVQVSGDITLPPVG 550  
Db 661 -----KTYVSDQG--TGWDLVQYQNKYGMFN 686  
Qy 551 ADPRKLVAFSDVLGSGQAHVVEVSATKVTWCNENLGRGRFGQPIITLPQFSOPATFNPAQV 610  
Db 687 PD-----GTIGKTLWL-----DYIQKRT 704  
Qy 611 YLADLDGSGPTDLI--YVHTNRLDIPLKNSGFAEPVTLRFPPEGLRFDHTCQLQOMADVQ 668  
Db 705 Y-----GDG-DDIVLPVINFRDLD--NKVGS-----AELNFRP----- 736  
Qy 669 GLGVASLILSVPHMSPHHRCDLTNMKPWLINEMNNMGVHHTLRYRSSQFWLDEKAAA 728  
Db 737 -----IKEHGLGATTIVSYGFANACDIDHLPQAQ 766  
Qy 729 LTTGQTVCYLPPFIHTLMQETDEISGNKLVTLRYARGAWDGRERFRGFGYVEQTD 788  
Db 767 ASNTQD--CY-----WQKWTPB-----GETDSKTGWFKKF--LVTVQV-- 799  
Qy 789 SHQLAQNAPERTPPALTKNWYATGLPVIDNALSTE-----YWRDDQAPAGSPRFTWQ 843  
Db 800 -----QVDPVTTN--QDGAFTVMTSYTYEDGAGWHFTNDPLIKDEDESMTDWR 846  
Qy 844 DNKDVPLT-----PEDDSRYWFNRAALKQLLSRLYGLDDSTNKHVPVTVTEPRSQVRL 899  
Db 847 GYQEVQVTTGAGAQKTKKSWLYRGLSG-----DRTSK----- 879  
Qy 900 QHTDSRYFVLWSVSVESRNYHYERIASDPQCSQNTLSSDRFGQPLKQLSVQYPRRQOPA 959  
Db 880 -----ADASATKTVTV--DDGDG----- 895  
Qy 960 INLYPDT--LPDQKLLANSY--DDQORQLRLTYQSSWHHLTNTVTVLGLPDSRSDIFTY 1016  
Db 896 -NNYTDSDLSGRILSTSLRDBDTCGTSHERTYHK--YWDH--NTAQYDGLPDPAR----FVR 946  
Qy 1017 GAENVPAGLNLLELSDKNSLIADDKPREYLGQOKTAYTDGONTT---PLQPTT--ROALI 1072  
Db 947 EKE-----TTNTKVSSGW--REHTVEYEDDTGEGASTTGLPMKTDWQSSV 993  
Qy 1073 AFTETTVF-----NOSTLSAFNGSIPSDKLS-----TTLEAGYQOQTNV----- 1111  
Db 994 SDNRCCTTYGRAYNTDNYDSTCAQRQRTVLQDQVHYSGVCSSSIADSKQDGYASTLYDNATS 1053  
Qy 1112 -----LPFRTGEDK--VWVAHHGYTDYGTAAQFWRPQKQSNQTLTG--KITL 1154  
Db 1054 VDANKPVDGNPTEGRTYTKSGSYRSTW-----SGYDDAGRV--MWSEGDGSHNRTLTKYSPAN 1108  
Qy 1155 IWDANYCVWVOTRDAAG-----LTTSAKYDWRPL--TPVQLTDINDNOHLITLDALGRPI 1207  
Db 1109 TWPLN--GVIVTTPDGDALPARTALTSTAMTSRFGKPTSIQDANGNVTKMSLDAAGRLV 1167

Qy 1208 TLRFW--GTENGKMTGYSSPE--KASFSPSPDVNAAIELKPLPVAQCQVYAPESWMPVLSQ 1265  
Db 1168 EV--WRPTETG-----SSPSMKFSYTIPTSTNSA----- 1194  
Qy 1266 KTFNRLAEQDWQKLYNARIITEDGRICTLAYRRVWVSQKAIPLQLISLNNGPRLPPHSL-- 1324  
Db 1195 -----GVPDAV-----DGYPHVATHVLQ 1212  
Qy 1325 TLTTDRYDHDPEQOIROQVVVSDGFGRLQAAARHEAGM--ARQRNEDGSLIINYQHTENR 1383  
Db 1213 SGITYLSSH-----AYVDGLGRARETQTPMGNGVDAATGNEVPNRQVSVTRYDSA 1262  
Qy 1384 WATGRTE--YDNKG-----QPIRTYQPYFLNDM--RYVSN----- 1415  
Db 1263 GNVGTGSAVFRNOCGTAGSGGFSKAKVEDLPSYTDLVL--DWAGRAITSRLQVNGASQDAGR 1321  
Qy 1416 -----DSARQKEAYADTHVYDPIGREIKVITAKGWFRRTLFTPMFTVNEDE---- 1462  
Db 1322 VDTTYDGDFTSVKNKVDAAADTYT--DVYGQVSKVVEHTG--SATYTTAYTYTAKDELIKI 1378  
Qy 1463 -----NDTA 1466  
Db 1379 TDPRGNDTS 1387

Search completed: December 16, 2005, 14:47:04  
Job time : 211 secs



*This Page Blank (uspio)*